PCT WORLD INTE Exp Mail EV335610938US
USAN 09/895,814 ƆN

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, C12N 15/10, 15/11

A2

(11) International Publication Number: WO 99/06550

(43) International Publication Date: 11 February 1999 (11.02.99)

(21) International Application Number: PCT/IB98/01232

(22) International Filing Date: 31 July 1998 (31.07.98)

(30) Priority Data:
08/905,144 1 August 1997 (01.08.97) US

(71) Applicant (*for all designated States except US*): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): DUMAS MILNE EDWARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire-de-Tours, F-75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). LACROIX, Bruno [FR/FR]; 93, route de Vourles, F-69230 Saint-Genis Laval (FR).

(74) Agents: MARTIN Jean-Jacques et al.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: 5' ESTs FOR SECRETED PROTEINS EXPRESSED IN PROSTATE

(57) Abstract

The sequences of 5' ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5' ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5' ESTs. The 5' ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5' ESTs. The 5' ESTs may also be used to design expression vectors and secretion vectors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

5' ESTs FOR SECRETED PROTEINS EXPRESSED IN PROSTATE

Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be isolated, sequenced, mapped, and characterized. Cloning vectors such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are able to accept DNA inserts ranging from 300 to 1000 kilobases (kb) or 100-400 kb in length respectively, thereby facilitating the manipulation and ordering of DNA sequences distributed over great distances on the human chromosomes. Automated DNA sequencing machines permit the rapid sequencing of human genes. Bioinformatics software enables the comparison of nucleic acid and protein sequences, thereby assisting in the characterization of human gene products.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bioinformatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bioinformatics software may mischaracterize the genomic sequences obtained. Thus, the software may produce false positives in which non-coding DNA is mischaracterized as coding DNA or false negatives in which coding DNA is mislabeled as non-coding DNA.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach,

sequencing is only performed on DNA which is derived from protein coding portions of the genome. Often, only short stretches of the cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then be used to isolate or purify extended cDNAs which include sequences adjacent to the EST sequences. The extended cDNAs may contain all of the sequence of the EST which was used to obtain them or only a portion of the sequence of the EST which was used to obtain them. In addition, the extended cDNAs may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the extended cDNAs may include portions of the coding sequence of the gene from which the EST was derived. It will be appreciated that there may be several extended cDNAs which include the EST sequence as a result of alternate splicing or the activity of alternative promoters.

In the past, these short EST sequences were often obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3′ untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3′ end of the mRNA is a result of the fact that typical techniques for obtaining cDNAs are not well suited for isolating cDNA sequences derived from the 5′ ends of mRNAs. (Adams *et al., Nature* 377:3-174, 1996; Hillier *et al., Genome Res.* 6:807-828, 1996).

In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include the full 5′ untranslated region of the mRNA from which the cDNA is derived. Such incomplete sequences may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5′ ends of mRNAs.

While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 50,000-100,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often

involved in cell to cell communication and may be responsible for producing a clinically relevant response in their target cells.

In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon-α, interferon-β, interferon-γ, and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney carcinoma, chemotherapy induced neutropenia and multiple sclerosis. For these reasons, extended cDNAs encoding secreted proteins or portions thereof represent a particularly valuable source of therapeutic agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences located at the 5′ ends of the coding sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the protein for which secretion is desired. In addition, portions of signal sequences may also be used to direct the intracellular import of a peptide or protein of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cell in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired protein must be selected. Thus, there exists a need to identify and characterize the 5′ portions of the genes for secretory proteins which encode signal peptides.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches

have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross, *et al., Nature Genetics* **6**: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock *et al., Genome Res.* **6**:327-335, 1996). Both of these approaches have their limits due to a lack of specificity or of comprehensiveness.

The present 5′ ESTs may be used to efficiently identify and isolate upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. (Theil, *BioFactors* **4**:87-93, 1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, ESTs containing the 5′ ends of secretory protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5′ coding sequences of genes encoding secretory proteins.

Summary of the Invention

The present invention relates to purified, isolated, or recombinant ESTs which include sequences derived from the authentic 5′ ends of their corresponding mRNAs. The term "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5′ EST. These sequences will be referred to hereinafter as "5′ ESTs." As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5′ EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^4-10^6 fold purification of the native message.

Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the 5′ EST is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the 5′ ESTs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched 5′ ESTs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched 5′ ESTs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched 5′ ESTs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Stringent", moderate," and "low" hybridization conditions are as defined in Example 29.

Unless otherwise indicated, a "complementary" sequence is fully complementary.

Thus, 5′ ESTs in cDNA libraries in which one or more 5′ ESTs make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant 5′ ESTs" as defined herein. Likewise, 5′ ESTs in a population of plasmids in which one or more 5′ EST of the present invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are " enriched recombinant 5′ ESTs" as defined herein. However, 5′ ESTs in cDNA libraries in which 5′ ESTs constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in

which backbone molecules having a 5′ EST insert are extremely rare, are not "enriched recombinant 5′ ESTs."

In particular, the present invention relates to 5′ ESTs which are derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

Such 5′ ESTs include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the genes from which the 5′ ESTs are derived. Generally, the signal peptides are located at the amino termini of secreted proteins.

Secreted proteins are translated by ribosomes associated with the "rough" endoplasmic reticulum. Generally, secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum. After delivery to the endoplasmic reticulum, secreted proteins may proceed through the Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across the cell membrane.

The 5′ ESTs of the present invention have several important applications. For example, they may be used to obtain and express cDNA clones which include the full protein coding sequences of the corresponding gene products, including the authentic translation start sites derived from the 5′ ends of the coding sequences of the mRNAs from which the 5′ ESTs are derived. These cDNAs will be referred to hereinafter as "full length cDNAs." These cDNAs may also include DNA derived from mRNA sequences upstream of the translation start site. The full length cDNA sequences may be used to express the proteins corresponding to the 5′ ESTs. As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating or

controlling a variety of human conditions. The 5′ ESTs may also be used to obtain the corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes the mRNA from which the 5′ EST was derived.

Alternatively, the 5′ ESTs may be used to obtain and express extended cDNAs encoding portions of the secreted protein. The portions may comprise the signal peptides of the secreted proteins or the mature proteins generated when the signal peptide is cleaved off. The portions may also comprise polypeptides having at least 10 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. Alternatively, the portions may comprise at least 15 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In some embodiments, the portions may comprise at least 25 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In other embodiments, the portions may comprise at least 40 amino acids encoded by the extended cDNAs or full length cDNAs.

Antibodies which specifically recognize the entire secreted proteins encoded by the extended cDNAs, full length cDNAs, or fragments thereof having at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 25 consecutive amino acids, or at least 40 consecutive amino acids may also be obtained as described below. Antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the signal peptides encoded by the extended cDNAs or full length cDNAs may also be obtained.

In some embodiments, the extended cDNAs obtained using the 5′ ESTs include the signal sequence. In other embodiments, the extended cDNAs obtained using the 5′ ESTs may include the full coding sequence for the mature protein (*i.e.* the protein generated when the signal polypeptide is cleaved off). In addition, the extended cDNAs obtained using the 5′ ESTs may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene expression.

As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the extended cDNAs or full length cDNAs obtained using the 5′ ESTs may be useful in treating or controlling a variety of human conditions.

The 5′ ESTs (or cDNAs or genomic DNAs obtained therefrom) may be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the 5′ ESTs. In addition, the present invention is useful for constructing a high resolution map of the human chromosomes.

The present invention also relates to secretion vectors capable of directing the secretion of a protein of interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the 5′ ESTs, such as promoters or upstream regulatory sequences.

Finally, the present invention may also be used for gene therapy to control or treat genetic diseases. Signal peptides may also be fused to heterologous proteins to direct their extracellular secretion.

Bacterial clones containing Bluescript plasmids having inserts containing the 5′ ESTs of the present invention (SEQ ID NOs: 38-315 are presently stored at 80°C in 4% (v/v) glycerol in the inventor's laboratories under the designations listed next to the SEQ ID NOs in II). The inserts may be recovered from the deposited materials by growing the appropriate clones on a suitable medium. The Bluescript DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5′ EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

One aspect of the present invention is a purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-315 or having a sequence complementary thereto. In one embodiment, the nucleic acid is recombinant.

Another aspect of the present invention is a purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary thereto.

Yet another aspect of the present invention is a purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-315 or one of the sequences complementary thereto. In one embodiment, the nucleic acid is recombinant.

A further aspect of the present invention is a purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315. In one embodiment, the nucleic acid is recombinant.

Another aspect of the present invention is a purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-315.

Still another aspect of the present invention is a method of making a cDNA encoding a human secretory protein, said human secretory protein being partially encoded by one of SEQ ID NOs 38-315, comprising the steps of contacting a collection of mRNA molecules from human cells with a primer comprising at least 15 consecutive nucleotides of a sequence complementary to one of SEQ ID NOs: 38-315; hybridizing said primer to an mRNA in said collection that encodes said protein; reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA; making a second cDNA strand complementary to said first cDNA strand; and isolating the resulting cDNA encoding said protein comprising said first cDNA strand and said second cDNA strand.

Another aspect of the invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the

cDNA comprises the full protein coding sequence of said protein which sequence is partially included in one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is a method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-315, comprising the steps of obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-315; contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-315 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA; identifying a cDNA which hybridizes to said detectable probe; and isolating said cDNA which hybridizes to said probe.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is a method of making a cDNA comprising one of the sequence of SEQ ID NOs: 38-315, comprising the steps of contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA; hybridizing said first primer to said polyA tail; reverse transcribing said mRNA to make a first cDNA strand; making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-315; and isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

In one embodiment of the method described in the two paragraphs above, the second cDNA strand is made by contacting said first cDNA strand with a first pair of primers, said

first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-315 and a third primer having a sequence therein which is included within the sequence of said first primer; performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product; contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs: 38-315 , and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and performing a second polymerase chain reaction, thereby generating a second PCR product.

One aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is the method described four paragraphs above in which the second cDNA strand is made by contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-315; hybridizing said second primer to said first strand cDNA; and extending said hybridized second primer to generate said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-315 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in of one of the sequences of SEQ ID NOs: 38-315.

Another aspect of the present invention is a method of making a protein comprising one of the sequences of SEQ ID NOs: 316-593, comprising the steps of obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NOs: 38-315; inserting said cDNA in an expression vector such that said cDNA is

operably linked to a promoter; introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and isolating said protein.

Another aspect of the present invention is an isolated protein obtainable by the method described in the preceding paragraph.

Another aspect of the present invention is a method of obtaining a promoter DNA comprising the steps of obtaining DNAs located upstream of the nucleic acids of SEQ ID NOs: 38-315 or the sequences complementary thereto; screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and isolating said DNA comprising said identified promoter. In one embodiment, the obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NOs: 38-315 or sequences complementary thereto. In another embodiment, the screening step comprises inserting said upstream sequences into a promoter reporter vector. In another embodiment, the screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

Another aspect of the present invention is an isolated promoter obtainable by the method described above.

Another aspect of the present invention is an isolated or purified protein comprising one of the sequences of SEQ ID NOs: 316-593.

Another aspect of the present invention is the inclusion of at least one of the sequences of SEQ ID NOs: 38-315, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315, or a fragment thereof of at least 15 consecutive nucleotides in an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length. In one embodiment, the array includes at least two of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides. In another embodiment, the array includes at least five of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides.

Another aspect of the present invention is a promoter having a sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 37.

Brief Description of the Drawings

Figure 1 is a summary of a procedure for obtaining cDNAs which have been selected to include the 5' ends of the mRNAs from which they derived.

Figure 2 shows the distribution of Von Heijne scores for 5' ESTs in each of the categories described herein and the probability that these 5' ESTs encode a signal peptide.

Figure 3 summarizes a general method used to clone and sequence extended cDNAs containing sequences adjacent to 5' ESTs.

Figure 4 (description of promoters structure isolated from SignalTag 5' ESTs) provides a schematic description of promoters isolated and the way they are assembled with the corresponding 5' tags.

Detailed Description of the Preferred Embodiment

Table IV is an analysis of the 43 amino acids located at the N terminus of all human SwissProt proteins to determine the frequency of false positives and false negatives using the techniques for signal peptide identification described herein.

Table V shows the distribution of 5' ESTs in each category described herein and the number of 5' ESTs in each category having a given minimum Von Heijne's score.

Table VI shows the distribution of 5' ESTs in each category described herein with respect to the tissue from which the 5' ESTs of the corresponding mRNA were obtained.

Table VII describes the transcription factor binding sites present in each of these promoters.

I. General Methods for Obtaining 5' ESTs derived from mRNAs with intact 5' ends

In order to obtain the 5' ESTs of the present invention, mRNAs with intact 5' ends must be obtained. Currently, there are two approaches for obtaining such mRNAs with intact 5' ends as described below: either chemical (1) or enzymatic (2).

1. Chemical Methods for Obtaining mRNAs having Intact 5' Ends

One of these approaches is a chemical modification method involving derivatization of the 5' ends of the mRNAs and selection of the derivatized mRNAs. The 5' ends of eukaryotic mRNAs possess a structure referred to as a "cap" which comprises a guanosine

methylated at the 7 position. The cap is joined to the first transcribed base of the mRNA by a 5′, 5′-triphosphate bond. In some instances, the 5′ guanosine is methylated in both the 2 and 7 positions. Rarely, the 5′ guanosine is trimethylated at the 2, 7 and 7 positions. In the chemical method for obtaining mRNAs having intact 5′ ends, the 5′ cap is specifically derivatized and coupled to a reactive group on an immobilizing substrate. This specific derivatization is based on the fact that only the ribose linked to the methylated guanosine at the 5′ end of the mRNA and the ribose linked to the base at the 3′ terminus of the mRNA, possess 2′, 3′-cis diols.

Optionally, the 2′, 3′-cis diol of the 3′ terminal ribose may be chemically modified, substituted, converted, or eliminated, leaving only the ribose linked to the methylated guanosine at the 5′ end of the mRNA with a 2′, 3′-cis diol. A variety of techniques are available for eliminating the 2′, 3′-cis diol on the 3′ terminal ribose. For example, controlled alkaline hydrolysis may be used to generate mRNA fragments in which the 3′ terminal ribose is a 3′-phosphate, 2′-phosphate or (2′, 3′)-cyclophosphate. Thereafter, the fragment which includes the original 3′ ribose may be eliminated from the mixture through chromatography on an oligodT column. Alternatively, a base which lacks the 2′, 3′-cis diol may be added to the 3′ end of the mRNA using an RNA ligase such as T4 RNA ligase. Example 1 below describes a method for ligation of a nucleoside diphosphate to the 3′ end of messenger RNA.

EXAMPLE 1

Ligation of the Nucleoside Diphosphate pCp to the 3' End of mRNA.

One μg of RNA was incubated in a final reaction medium of 10 μl in the presence of 5 U of T_4 phage RNA ligase in the buffer provided by the manufacturer (Gibco - BRL), 40 U of the RNase inhibitor RNasin (Promega) and, 2 μl of ^{32}pCp (Amersham #PB 10208). The incubation was performed at 37°C for 2 hours or overnight at 7-8°C.

Following modification or elimination of the 2′, 3′-cis diol at the 3′ ribose, the 2′, 3′-cis diol present at the 5′ end of the mRNA may be oxidized using reagents such as $NaBH_4$, $NaBH_3CN$, or sodium periodate, thereby converting the 2′, 3′-cis diol to a dialdehyde.

Example 2 describes the oxidation of the 2′, 3′-cis diol at the 5′ end of the mRNA with sodium periodate.

EXAMPLE 2

Oxidation of 2′, 3′-cis diol at the 5′ End of the mRNA with Sodium Periodate

0.1 OD unit of either a capped oligoribonucleotide of 47 nucleotides (including the cap) or an uncapped oligoribonucleotide of 46 nucleotides were treated as follows. The oligoribonucleotides were produced by *in vitro* transcription using the transcription kit "AmpliScribe T7" (Epicentre Technologies). As indicated below, the DNA template for the RNA transcript contained a single cytosine. To synthesize the uncapped RNA, all four NTPs were included in the *in vitro* transcription reaction. To obtain the capped RNA, GTP was replaced by an analogue of the cap, m7G(5′)ppp(5′)G. This compound, recognized by the polymerase, was incorporated into the 5′ end of the nascent transcript during the initiation of transcription but was not incorporated during the extension step. Consequently, the resulting RNA contained a cap at its 5′ end. The sequences of the oligoribonucleotides produced by the *in vitro* transcription reaction were:

+Cap:

5′m7GpppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3′ (SEQ ID NO:1)

-Cap:

5′-pppGCAUCCUACUCCCAUCCAAUUCCACCCUAACUCCUCCCAUCUCCAC-3′ (SEQ ID NO:2)

The oligoribonucleotides were dissolved in 9 µl of acetate buffer (0.1 M sodium acetate, pH 5.2) and 3 µl of freshly prepared 0.1 M sodium periodate solution. The mixture was incubated for 1 hour in the dark at 4°C or room temperature. Thereafter, the reaction was stopped by adding 4 µl of 10% ethylene glycol. The product was ethanol precipitated, resuspended in at least 10 µl of water or appropriate buffer and dialyzed against water.

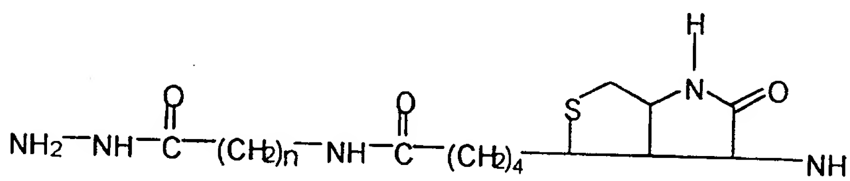
The resulting aldehyde groups may then be coupled to molecules having a reactive amine group, such as hydrazine, carbazide, thiocarbazide or semicarbazide groups, in order to facilitate enrichment of the 5′ ends of the mRNAs. Molecules having

reactive amine groups which are suitable for use in selecting mRNAs having intact 5′ ends include avidin, proteins, antibodies, vitamins, ligands capable of specifically binding to receptor molecules, or oligonucleotides. Example 3 below describes the coupling of the resulting dialdehyde to biotin.

EXAMPLE 3

Coupling of the Dialdehyde at the 5' End of Transcripts with Biotin

The oxidation product obtained in Example 2 was dissolved in 50 μl of sodium acetate at a pH between 5 and 5.2 and 50 μl of freshly prepared 0.02 M solution of biotin hydrazide in a methoxyethanol/water mixture (1:1) of formula:



In the compound used in these experiments, n=5. However, it will be appreciated that other commercially available hydrazides may also be used, such as molecules of the above formula in which n varies from 0 to 5. The mixture was then incubated for 2 hours at 37°C, precipitated with ethanol and dialyzed against distilled water. Example 4 demonstrates the specificity of the biotinylation reaction.

EXAMPLE 4

Specificity of Biotinylation of Capped Transcripts

The specificity of the biotinylation for capped mRNAs was evaluated by gel electrophoresis of the following samples:

Sample 1. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2 and labeled with ^{32}pCp as described in Example 1.

Sample 2. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2, labeled with ^{32}pCp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Sample 3. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2 and labeled with ^{32}pCp as described in Example 1.

Sample 4. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2, labeled with ^{32}pCp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Samples 1 and 2 had identical migration rates, demonstrating that the uncapped RNAs were not oxidized and biotinylated. Sample 3 migrated more slowly than Samples 1 and 2, while Sample 4 exhibited the slowest migration. The difference in migration of the RNAs in Samples 3 and 4 demonstrates that the capped RNAs were specifically biotinylated.

In some cases, mRNAs having intact 5′ ends may be enriched by binding the molecule containing a reactive amine group to a suitable solid phase substrate such as the inside of the vessel containing the mRNAs, magnetic beads, chromatography matrices, or nylon or nitrocellulose membranes. For example, where the molecule having a reactive amine group is biotin, the solid phase substrate may be coupled to avidin or streptavidin. Alternatively, where the molecule having the reactive amine group is an antibody or receptor ligand, the solid phase substrate may be coupled to the cognate antigen or receptor. Finally, where the molecule having a reactive amine group comprises an oligonucleotide, the solid phase substrate may comprise a complementary oligonucleotide.

The mRNAs having intact 5′ ends may be released from the solid phase following the enrichment procedure. For example, where the dialdehyde is coupled to biotin hydrazide and the solid phase comprises streptavidin, the mRNAs may be released from the solid phase by simply heating to 95 degrees Celsius in 2% SDS. In some methods, the molecule having a reactive amine group may also be cleaved from the mRNAs having intact 5′ ends following enrichment. Example 5 describes the capture of biotinylated mRNAs with streptavidin coated beads and the release of the biotinylated mRNAs from the beads following enrichment.

EXAMPLE 5

Capture and Release of Biotinylated mRNAs Using Streptavidin Coated Beads

The streptavidin coated magnetic beads were prepared according to the manufacturer's instructions (CPG Inc., USA). The biotinylated mRNAs were added to a

hybridization buffer (1.5 M NaCl, pH 5 - 6). After incubating for 30 minutes, the unbound and nonbiotinylated material was removed. The beads were then washed several times in water with 1% SDS. The beads thus obtained were incubated for 15 minutes at 95°C in water containing 2% SDS.

Example 6 demonstrates the efficiency with which biotinylated mRNAs were recovered from the streptavidin coated beads.

EXAMPLE 6

Efficiency of Recovery of Biotinylated mRNAs

The efficiency of the recovery procedure was evaluated as follows. Capped RNAs were labeled with ^{32}pCp, oxidized, biotinylated and bound to streptavidin coated beads as described above. Subsequently, the bound RNAs were incubated for 5, 15 or 30 minutes at 95°C in the presence of 2% SDS.

The products of the reaction were analyzed by electrophoresis on 12% polyacrylamide gels under denaturing conditions (7 M urea). The gels were subjected to autoradiography. During this manipulation, the hydrazone bonds were not reduced.

Increasing amounts of nucleic acids were recovered as incubation times in 2% SDS increased, demonstrating that biotinylated mRNAs were efficiently recovered.

In an alternative method for obtaining mRNAs having intact 5′ ends, an oligonucleotide which has been derivatized to contain a reactive amine group is specifically coupled to mRNAs having an intact cap. Preferably, the 3′ end of the mRNA is blocked prior to the step in which the aldehyde groups are joined to the derivatized oligonucleotide, as described above, so as to prevent the derivatized oligonucleotide from being joined to the 3′ end of the mRNA. For example, pCp may be attached to the 3′ end of the mRNA using T4 RNA ligase as described in example 1. However, as discussed above, blocking the 3′ end of the mRNA is an optional step. Derivatized oligonucleotides may be prepared as described in Example 7.

EXAMPLE 7

Derivatization of Oligonucleotides

An oligonucleotide phosphorylated at its 3′ end was converted to a 3′ hydrazide in 3′ by treatment with an aqueous solution of hydrazine or of dihydrazide of the formula $H_2N(R1)NH_2$ at about 1 to 3 M, and at pH 4.5 at a temperature of 8°C overnight. This incubation was performed in the presence of a carbodiimide type agent soluble in water such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide at a final concentration of 0.3 M.

The derivatized oligonucleotide was then separated from the other agents and products using a standard technique for isolating oligonucleotides.

As discussed above, the mRNAs to be enriched may be treated to eliminate the 3′ OH groups which may be present thereon. This may be accomplished by enzymatic ligation of sequences lacking a 3′ OH, such as pCp, as described in Example 1. Alternatively, the 3′ OH groups may be eliminated by alkaline hydrolysis as described in Example 8 below.

EXAMPLE 8

Elimination of 3' OH Groups of mRNA Using Alkaline Hydrolysis

In a total volume of 100 μl of 0.1 N sodium hydroxide, 1.5 μg mRNA is incubated for 40 to 60 minutes at 4°C. The solution is neutralized with acetic acid and precipitated with ethanol.

Following the optional elimination of the 3′ OH groups, the diol groups at the 5′ ends of the mRNAs are oxidized as described below in Example 9.

EXAMPLE 9

Oxidation of Diols of mRNA

Up to 1 OD unit of RNA was dissolved in 9 μl of buffer (0.1 M sodium acetate, pH 6-7) or water and 3 μl of freshly prepared 0.1 M sodium periodate solution. The reaction was incubated for 1 h in the dark at 4°C or room temperature. Following the incubation, the reaction was stopped by adding 4 μl of 10% ethylene glycol. Thereafter the mixture was incubated at room temperature for 15 minutes. After ethanol precipitation, the product was resuspended in at least 10 μl of water or appropriate buffer and dialyzed against water.

Following oxidation of the diol groups at the 5′ ends of the mRNAs, the derivatized oligonucleotide was joined to the resulting aldehydes as described in Example 10.

EXAMPLE 10

Ligature of Aldehydes of mRNA to Derivatized Oligonucleotides

The oxidized mRNA was dissolved in an acidic medium such as 50 μl of sodium acetate pH 4-6. Fifty μl of a solution of the derivatized oligonucleotide were added in order to obtain an mRNA:derivatized oligonucleotide ratio of 1:20. The mixture was reduced with a borohydride and incubated for 2 h at 37°C or overnight (14 h) at 10°C. The mixture was then ethanol precipitated, resuspended in 10 μl or more of water or appropriate buffer and dialyzed against distilled water. If desired, the resulting product may be analyzed using acrylamide gel electrophoresis, HPLC analysis, or other conventional techniques.

Following the attachment of the derivatized oligonucleotide to the mRNAs, a reverse transcription reaction may be performed as described in Example 11 below.

EXAMPLE 11

Reverse Transcription of mRNAs Ligatured to Derivatized Oligonucleotides

An oligodeoxyribonucleotide was derivatized as follows. Three OD units of an oligodeoxyribonucleotide of sequence 5'ATCAAGAATTCGCACGAGACCATTA3' (SEQ ID NO:3) having 5′-OH and 3′-P ends were dissolved in 70 μl of a 1.5 M hydroxybenzotriazole solution, pH 5.3, prepared in dimethylformamide/water (75:25) containing 2 μg of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. The mixture was incubated for 2 h 30 min at 22°C and then precipitated twice in $LiClO_4$/acetone. The pellet was resuspended in 200 μl of 0.25 M hydrazine and incubated at 8°C from 3 to 14 h. Following the hydrazine reaction, the mixture was precipitated twice in $LiClO_4$/acetone.

The messenger RNAs to be reverse transcribed were extracted from blocks of placenta having sides of 2 cm which had been stored at -80°C. The total RNA was extracted using conventional acidic phenol techniques. Oligo-dT chromatography was used to purify the mRNAs. The integrity of the mRNAs was checked by Northern-blotting.

The diol groups on 7 μg of the placental mRNAs were oxidized as described above in Example 9. The derivatized oligonucleotide was joined to the mRNAs as described in Example 10 above except that the precipitation step was replaced by an exclusion chromatography step to remove derivatized oligodeoxyribonucleotides which were not joined to mRNAs. Exclusion chromatography was performed as follows:

Ten ml of Ultrogel AcA34 (BioSepra#230151) gel, a mix of agarose and acrylamide, were equilibrated in 50 ml of a solution of 10 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA, and 0.05% SDS. The mixture was allowed to sediment. The supernatant was eliminated and the gel was resuspended in 50 ml of buffer. This procedure was repeated 2 or 3 times.

A glass bead (diameter 3 mm) was introduced into a 2 ml disposable pipette (length 25 cm). The pipette was filled with the gel suspension until the height of the gel stabilized at 1 cm from the top of the pipette. The column was then equilibrated with 20 ml of equilibration buffer (10 mM Tris HCl pH 7.4, 20 mM NaCl).

Ten μl of the mRNA which had reacted with the derivatized oligonucleotide were mixed in 39 μl of 10 mM urea and 2 μl of blue-glycerol buffer, which had been prepared by dissolving 5 mg of bromophenol blue in 60% glycerol (v/v), and passing the mixture through a 0.45 μm diameter filter.

The column was then loaded with the mRNAs coupled to the oligonucleotide. As soon as the sample had penetrated, equilibration buffer was added. Hundred μl fractions were then collected. Derivatized oligonucleotide which had not been attached to mRNA appeared in fraction 16 and later fractions. Thus, fractions 3 to 15 were combined and precipitated with ethanol.

To determine whether the derivatized oligonucleotide was actually linked to mRNA, one tenth of the combined fractions were spotted twice on a nylon membrane and hybridized to a radioactive probe using conventional techniques. The ^{32}P labeled probe used in these hybridizations was an oligodeoxyribonucleotide of sequence 5'TAATGGTCTCGTGCGAATTCTTGAT3' (SEQ ID NO:4) anticomplementary to the derivatized oligonucleotide. A signal observed after autoradiography, indicated that the derivatized oligonucleotide had been truly joined to the mRNA.

The remaining nine tenth of the mRNAs which had reacted with the derivatized oligonucleotide was reverse transcribed as follows. A reverse transcription reaction was

carried out with reverse transcriptase following the manufacturer's instructions and 50 pmol of nonamers with random sequence as primers.

To ensure that reverse transcription had been carried out through the cap structure, two types of experiments were performed.

In the first approach, after elimination of RNA of the cDNA:RNA heteroduplexes obtained from the reverse transcription reaction by an alkaline hydrolysis, a portion of the resulting single stranded cDNAs was spotted on a positively charged membrane and hybridized, using conventional methods, to a ^{32}P labeled probe having a sequence identical to that of the derivatized oligonucleotide. Control spots containing, 1 pmol, 100 fmol, 50 fmol, 10 fmol and 1 fmol of a control oligodeoxyribonucleotide of sequence identical to that of the derivatized oligonucleotide were included. The signal observed in the spots containing the cDNA indicated that approximately 15 fmol of the derivatized oligonucleotide had been reverse transcribed. These results demonstrate that the reverse transcription can be performed through the cap and, in particular, that reverse transcriptase crosses the 5'-P-P-P-5' bond of the cap of eukaryotic messenger RNAs.

In the second type of experiment, the single stranded cDNAs obtained from the above first strand synthesis were used as template for PCR reactions. Two types of reactions were carried out. First, specific amplification of the mRNAs for alpha globin, dehydrogenase, pp15 and elongation factor E4 were carried out using the following pairs of oligodeoxyribonucleotide primers.

alpha-globin

GLO-S: 5'CCG ACA AGA CCA ACG TCA AGG CCG C3' (SEQ ID NO:5)

GLO-As: 5'TCA CCA GCA GGC AGT GGC TTA GGA G 3' (SEQ ID NO:6)

dehydrogenase

3 DH-S: 5'AGT GAT TCC TGC TAC TTT GGA TGG C3' (SEQ ID NO:7)

3 DH-As: 5'GCT TGG TCT TGT TCT GGA GTT TAG A3' (SEQ ID NO:8)

pp15

PP15-S: 5'TCC AGA ATG GGA GAC AAG CCA ATT T3' (SEQ ID NO:9)

PP15-As: 5'AGG GAG GAG GAA ACA GCG TGA GTC C3' (SEQ ID NO:10)

Elongation factor E4

EFA1-S: 5'ATG GGA AAG GAA AAG ACT CAT ATC A3' (SEQ ID NO:11)

EF1A-As: 5'AGC AGC AAC AAT CAG GAC AGC ACA G3' (SEQ ID NO:12)

Second, non specific amplifications were also carried out with the antisense oligodeoxyribonucleotides of the pairs described above and with a primer derived from the sequence of the derivatized oligodeoxyribonucleotide (5'ATCAAGAATTCGCACGAGACCATTA3') (SEQ ID NO:13).

One twentieth of the following RT-PCR product samples were run on a 1.5% agarose gel and stained with ethidium bromide.

Sample 1: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the presence of cDNA.

Sample 2: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the absence of added cDNA.

Sample 3: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the presence of cDNA.

Sample 4: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the absence of added cDNA.

Sample 5: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the presence of cDNA.

Sample 6: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the absence of added cDNA.

Sample 7: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the presence of added cDNA.

Sample 8: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the absence of added cDNA.

A band of the size expected for the PCR product was observed only in samples 1, 3, 5 and 7, thus indicating the presence of the corresponding sequence in the cDNA population.

PCR reactions were also carried out with the antisense oligonucleotides of the globin and dehydrogenase primers (SEQ ID NOs 6 and 8) and an oligonucleotide whose sequence corresponds to that of the derivatized oligonucleotide. The presence of PCR products of the expected size in the samples equivalent to above samples 1 and 3 indicated that the derivatized oligonucleotide had been linked to mRNA.

The above examples summarize the chemical procedure for enriching mRNAs for those having intact 5′ ends as illustrated in Figure 1. Further detail regarding the chemical approaches for obtaining such mRNAs are disclosed in International Application No. WO96/34981, published November 7, 1996, which is incorporated herein by reference. Strategies based on the above chemical modifications to the 5′ cap structure may be utilized to generate cDNAs selected to include the 5′ ends of the mRNAs from which they derived. In one version of such procedures, the 5′ ends of the mRNAs are modified as described above. Thereafter, a reverse transcription reaction is conducted to extend a primer complementary to the 5′ end of the mRNA. Single stranded RNAs are eliminated to obtain a population of cDNA/mRNA heteroduplexes in which the mRNA includes an intact 5′ end. The resulting heteroduplexes may be captured on a solid phase coated with a molecule capable of interacting with the molecule used to derivatize the 5′ end of the mRNA. Thereafter, the strands of the heteroduplexes are separated to recover single stranded first cDNA strands which include the 5′ end of the mRNA. Second strand cDNA synthesis may then proceed using conventional techniques. For example, the procedures disclosed in WO 96/34981 or in Carninci. *et al.*, *Genomics* 37:327-336, 1996, the disclosures of which are incorporated herein by reference, may be employed to select cDNAs which include the sequence derived from the 5′ end of the coding sequence of the mRNA.

Following ligation of the oligonucleotide tag to the 5′ cap of the mRNA, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5′ end of the mRNA. Following elimination of the RNA component of the resulting heteroduplex using standard techniques, second strand cDNA synthesis is conducted with a primer complementary to the oligonucleotide tag.

2. Enzymatic Methods for Obtaining mRNAs having Intact 5′ Ends

Other techniques for selecting cDNAs extending to the 5′ end of the mRNA from which they are derived are fully enzymatic. Some versions of these techniques are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al., Gene* **150**:243-250, 1994, the disclosures of which are incorporated herein by reference.

Briefly, in such approaches, isolated mRNA is treated with alkaline phosphatase to remove the phosphate groups present on the 5′ ends of uncapped incomplete mRNAs. Following this procedure, the cap present on full length mRNAs is enzymatically removed with a decapping enzyme such as T4 polynucleotide kinase or tobacco acid pyrophosphatase. An oligonucleotide, which may be either a DNA oligonucleotide or a DNA-RNA hybrid oligonucleotide having RNA at its 3′ end, is then ligated to the phosphate present at the 5′ end of the decapped mRNA using T4 RNA ligase. The oligonucleotide may include a restriction site to facilitate cloning of the cDNAs following their synthesis. Example 12 below describes one enzymatic method based on the doctoral thesis of Dumas.

EXAMPLE 12

Enzymatic Approach for Obtaining 5′ ESTs

Twenty micrograms of PolyA+ RNA were dephosphorylated using Calf Intestinal Phosphatase (Biolabs). After a phenol chloroform extraction, the cap structure of mRNA was hydrolysed using the Tobacco Acid Pyrophosphatase (purified as described by Shinshi *et al.*., *Biochemistry* 15: 2185-2190, 1976) and a hemi 5′DNA/RNA-3′ oligonucleotide having an unphosphorylated 5′ end, a stretch of adenosine ribophosphate at the 3′ end, and an EcoRI site near the 5′ end was ligated to the 5′P ends of mRNA using the T4 RNA ligase (Biolabs). Oligonucleotides suitable for use in this procedure are preferably 30 to 50 bases in length. Oligonucleotides having an unphosphorylated 5′ end may be synthesized by adding a fluorochrome at the 5′ end. The inclusion of a stretch of adenosine ribophosphates at the 3′ end of the oligonucleotide increases ligation efficiency. It will be appreciated that the oligonucleotide may contain cloning sites other than EcoRI.

Following ligation of the oligonucleotide to the phosphate present at the 5′ end of the decapped mRNA, first and second strand cDNA synthesis is carried out using conventional methods or those specified in EP0 625,572 and Kato *et al. supra*, and Dumas Milne Edwards, *supra*, the disclosures of which are incorporated herein by reference. The resulting cDNA may then be ligated into vectors such as those disclosed in Kato *et al.*, *supra* or other nucleic acid vectors known to those skilled in the art using techniques such as those described in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference.

II. Obtention and Characterization of the 5′ ESTs of the Present Invention

The 5' ESTs of the present invention were obtained using the aforementioned chemical and enzymatic approaches for enriching mRNAs for those having intact 5′ ends as decribed below.

1. Obtention of 5' ESTS Using mRNAs with Intact 5' Ends

First, mRNAs were prepared as described in Example 13 below.

EXAMPLE 13

Preparation of mRNA With Intact 5' Ends

Total human RNAs or polyA$^+$ RNAs derived from 29 different tissues were respectively purchased from LABIMO and CLONTECH and used to generate 44 cDNA libraries as follows. The purchased RNA had been isolated from cells or tissues using acid guanidium thiocyanate-phenol-chloroform extraction (Chomczyniski and Sacchi, *Analytical Biochemistry* **162**:156-159, 1987). PolyA$^+$ RNA was isolated from total RNA (LABIMO) by two passes of oligo dT chromatography, as described by Aviv and Leder, *Proc. Natl. Acad. Sci. USA* **69**:1408-1412, 1972 in order to eliminate ribosomal RNA.

The quality and the integrity of the polyA+ RNAs were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded. Contamination of the polyA$^+$ mRNAs by ribosomal sequences was checked using Northern blots and a probe derived from the sequence of the 28S rRNA. Preparations of mRNAs with

less than 5% of rRNAs were used in library construction. To avoid constructing libraries with RNAs contaminated by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly expressed fungal mRNAs was examined using PCR.

Following preparation of the mRNAs, the above described chemical and/or the enzymatic procedures for enriching mRNAs for thoses having intact 5′ ends were employed to obtain 5′ ESTs from various tissues. In both approaches, an oligonucleotide tag was attached to the 5′ ends of the mRNAs. The oligonucleotide tag had an EcoRI site therein to facilitate later cloning procedures. To facilitate the processing of single stranded and double stranded cDNA obtained in the construction of the librairies, the same nucleotidic sequence was used to design the ligated oligonucleotide in both chemical and enzymatic approaches. Nevertheless, in the chemical procedure, the tag used was an oligodeoxyribonucleotide which was linked to the cap of the mRNA whereas in the enzymatic ligation, the tag was a chimeric hemi 5'DNA/RNA3' oligonucleotide which was ligated to the 5' end of decapped mRNA as described in example 12.

Following attachment of the oligonucleotide tag to the mRNA by either the chemical or enzymatic methods, the integrity of the mRNA was examined by performing a Northern blot with 200 to 500 ng of mRNA using a probe complementary to the oligonucleotide tag before performing the first strand synthesis as described in example 14.

EXAMPLE 14

cDNA Synthesis Using mRNA Templates Having Intact 5′ Ends

For the mRNAs joined to oligonucleotide tags using both the chemical and enzymatic methods, first strand cDNA synthesis was performed using the Superscript II (Gibco BRL) or the Rnase H Minus M-MLV (Promega) reverse transcriptase with random nonamers as primers. In order to protect internal EcoRI sites in the cDNA from digestion at later steps in the procedure, methylated dCTP was used for first strand synthesis. After removal of RNA by an alkaline hydrolysis, the first strand of cDNA was precipitated using isopropanol in order to eliminate residual primers.

For both the chemical and the enzymatic methods, the second strand of the cDNA was synthesized with a Klenow fragment using a primer corresponding to the 5′ end of the

ligated oligonucleotide described in Example 12. Preferably, the primer is 20-25 bases in length. Methylated dCTP was also used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following cDNA synthesis, the cDNAs were cloned into pBlueScript as described in Example 15 below.

EXAMPLE 15

Cloning of cDNAsderived from mRNA with intact 5' ends into BlueScript

Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA polymerase (Biolabs) and the cDNA was digested with EcoRI. Since methylated dCTP was used during cDNA synthesis, the EcoRI site present in the tag was the only hemi-methylated site, hence the only site susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra) and fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was directionally cloned into the SmaI and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were then selected as described in Example 16 below.

EXAMPLE 16

Selection of Clones Having the Oligonucleotide Tag Attached Thereto

The plasmid DNAs containing 5′ EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows. Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry *et al.*, *Biotechniques*, 13: 124-131, 1992. In this procedure, the single stranded DNA was hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3′ end of the oligonucleotide described in Example 13. Preferably, the primer has a length of 20-25

bases. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. Alternatively, protocoles such as the one described in the Gene Trapper kit available from Gibco BRL may be used. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5′ tag oligonucleotide was estimated to typically rank between 90 and 98% using dot blot analysis.

Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

EXAMPLE 17

Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer, Applied Biosystems Division, Foster City, CA), using standard SETA-A and SETA-B primers (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer). Sequencing reactions were performed using PE 9600 thermocyclers with standard dye-primer chemistry and ThermoSequenase (Amersham Pharmacia Biotech). The primers used were either T7 or 21M13 (available from Genset SA) as appropriate. The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with ethanol, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

2. Computer analysis of the Obtained 5' ESTs: Construction of NetGene and SignalTag databases

The sequence data from the 44 cDNA libraries made as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller, working using a Unix system, automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector or ligation oligonucleotides were automatically removed from the EST sequences. However, the resulting EST sequences may contain 1 to 5 bases belonging to the above mentioned sequences at their 5′ end. If needed, these can easily be removed on a case to case basis.

Following sequencing as described above, the sequences of the 5′ ESTs were entered in NetGene™, a proprietary database called for storage and manipulation as described below. It will be appreciated by those skilled in the art that the data could be stored and manipulated on any medium which can be read and accessed by a computer. Computer readable media include magnetically, optically, or electronically readable media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, RAM, or ROM as well as other types of other media known to those skilled in the art.

In addition, the sequence data may be stored and manipulated in a variety of data processor programs in a diversity of formats. For instance, the sequence data may be stored as text in a word processing file, such as Microsoft WORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE.

The computer readable media on which the sequence information is stored may be in a personal computer, a network, a server or other computer systems known to those skilled in the art. The computer or other system preferably includes the storage media described above, and a processor for accessing and manipulating the sequence data. Once the sequence data has been stored, it may be manipulated and searched to locate those stored sequences which contain a desired nucleic acid sequence or which encode a protein having a particular functional domain. For example, the stored sequence information may be compared to other

known sequences to identify homologies, motifs implicated in biological function, or structural motifs.

Programs which may be used to search or compare the stored sequences include the MacPattern (EMBL), BLAST, and BLAST2 program series (NCBI), basic local alignment search tool programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul *et al*, *J. Mol. Biol.* **215**: 403, 1990) and FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* **85**: 2444, 1988). The BLAST programs then extend the alignments on the basis of defined match and mismatch criteria.

Motifs which may be detected using the above programs and those described in Example 28 include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Before searching the cDNAs in the NetGene™ database for sequence motifs of interest, cDNAs derived from mRNAs which were not of interest were identified and eliminated from further consideration as described in Example 18 below.

EXAMPLE 18

Elimination of Undesired Sequences from Further Consideration

5′ ESTs in the NetGene™ database which were derived from undesired sequences such as transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs, fungal RNAs, Alu sequences, L1 sequences, or repeat sequences were identified using the FASTA and BLASTN programs with the parameters listed in Table I.

To eliminate 5′ ESTs encoding tRNAs from further consideration, the 5′ EST sequences were compared to the sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. The comparison was performed using FASTA on both strands of the 5′ ESTs. Sequences having more than 80% homology over more than 60 nucleotides were identified as tRNA. Of the 144,341 sequences screened, 26 were identified as tRNAs and eliminated from further consideration.

To eliminate 5′ ESTs encoding rRNAs from further consideration, the 5′ EST sequences were compared to the sequences of 2497 known rRNAs obtained from EMBL release 38, of which 73 were human. The comparison was performed using BLASTN on both strands of the 5′ ESTs with the parameter S=108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as rRNAs. Of the 144,341 sequences screened, 3,312 were identified as rRNAs and eliminated from further consideration.

To eliminate 5′ ESTs encoding mtRNAs from further consideration, the 5′ EST sequences were compared to the sequences of the two known mitochondrial genomes for which the entire genomic sequences are available and all sequences transcribed from these mitochondrial genomes including tRNAs, rRNAs, and mRNAs for a total of 38 sequences. The comparison was performed using BLASTN on both strands of the 5′ ESTs with the parameter S=108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as mtRNAs. Of the 144,341 sequences screened, 6,110 were identified as mtRNAs and eliminated from further consideration.

Sequences which might have resulted from exogenous contaminants were eliminated from further consideration by comparing the 5′ EST sequences to release 46 of the EMBL bacterial and fungal divisions using BLASTN with the parameter S=144. All sequences having more than 90% homology over at least 40 nucleotides were identified as exogenous contaminants. Of the 42 cDNA libraries examined, the average percentages of prokaryotic and fungal sequences contained therein were 0.2% and 0.5% respectively. Among these sequences, only one could be identified as a sequence specific to fungi. The others were either fungal or prokaryotic sequences having homologies with vertebrate sequences or including repeat sequences which had not been masked during the electronic comparison.

In addition, the 5′ ESTs were compared to 6093 Alu sequences and 1115 L1 sequences to mask 5′ ESTs containing such repeat sequences. 5′ ESTs including THE and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats were also eliminated from further consideration. On average, 11.5% of the sequences in the libraries contained repeat sequences. Of this 11.5%, 7% contained Alu repeats, 3.3% contained L1 repeats and the remaining 1.2% were derived from the other screened types of repetitive sequences. These percentages are consistent with those found in cDNA libraries prepared by

other groups. For example, the cDNA libraries of Adams *et al.* contained between 0% and 7.4% Alu repeats depending on the source of the RNA which was used to prepare the cDNA library (Adams *et al.*, *Nature* 377:174, 1996).

The sequences of those 5′ ESTs remaining after the elimination of undesirable sequences were compared with the sequences of known human mRNAs to determine the accuracy of the sequencing procedures described above.

EXAMPLE 19

Measurement of Sequencing Accuracy by Comparison to Known Sequences

To further determine the accuracy of the sequencing procedure described above, the sequences of 5′ ESTs derived from known sequences were identified and compared to the original known sequences. First, a FASTA analysis with overhangs shorter than 5 bp on both ends was conducted on the 5′ ESTs to identify those matching an entry in the public human mRNA database. The 6655 5′ ESTs which matched a known human mRNA were then realigned with their cognate mRNA and dynamic programming was used to include substitutions, insertions, and deletions in the list of "errors" which would be recognized. Errors occurring in the last 10 bases of the 5′ EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy.

This analysis revealed that the sequences incorporated in the NetGene™ database had an accuracy of more than 99.5%.

To determine the efficiency with which the above selection procedures select cDNAs which include the 5′ ends of their corresponding mRNAs, the following analysis was performed.

EXAMPLE 20

Determination of Efficiency of 5′ EST Selection

To determine the efficiency at which the above selection procedures isolated 5′ ESTs which included sequences close to the 5′ end of the mRNAs from which they derived, the sequences of the ends of the 5′ ESTs derived from the elongation factor 1 subunit α and

ferritin heavy chain genes were compared to the known cDNA sequences of these genes. Since the transcription start sites of both genes are well characterized, they may be used to determine the percentage of derived 5′ ESTs which included the authentic transcription start sites.

For both genes, more than 95% of the obtained 5' ESTs actually included sequences close to or upstream of the 5′ end of the corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5′ ESTs from ESTs in the NetGene™ database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for comparison. The 5' ends of more than 85% of 5′ ESTs derived from mRNAs included in the GeneBank database were located close to the 5′ ends of the known sequence. As some of the mRNA sequences available in the GenBank database are deduced from genomic sequences, a 5′ end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5′ ends of their corresponding mRNAs.

The EST libraries made above included multiple 5′ ESTs derived from the same mRNA. The sequences of such 5′ ESTs were compared to one another and the longest 5′ ESTs for each mRNA were identified. Overlapping cDNAs were assembled into continuous sequences (contigs). The resulting continuous sequences were then compared to public databases to gauge their similarity to known sequences, as described in Example 21 below.

EXAMPLE 21

Clustering of the 5′ ESTs and Calculation of Novelty Indices for cDNA Libraries

For each sequenced EST library, the sequences were clustered by the 5′ end. Each sequence in the library was compared to the others with BLASTN2 (direct strand, parameters S=107). ESTs with High Scoring Segment Pairs (HSPs) at least 25 bp long, having 95% identical bases and beginning closer than 10 bp from each EST 5′ end were grouped. The longest sequence found in the cluster was used as representative of the group. A global clustering between libraries was then performed leading to the definition of super-contigs.

To assess the yield of new sequences within the EST libraries, a novelty rate (NR) was defined as: NR= 100 X (Number of new unique sequences found in the library/Total number of sequences from the library). Typically, novelty rating ranged between 10% and 41% depending on the tissue from which the EST library was obtained. For most of the libraries, the random sequencing of 5′ EST libraries was pursued until the novelty rate reached 20%.

Following characterization as described above, the collection of 5′ ESTs in NetGene™ was screened to identify those 5′ ESTs bearing potential signal sequences as described in Example 22 below.

EXAMPLE 22

Identification of Potential Signal Sequences in 5′ ESTs

The 5′ ESTs in the NetGene™ database were screened to identify those having an uninterrupted open reading frame (ORF) longer than 45 nucleotides beginning with an ATG codon and extending to the end of the EST. Approximately half of the cDNA sequences in NetGene™ contained such an ORF. The ORFs of these 5′ ESTs were then searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, *Nucleic Acids Res.* **14**:4683-4690, 1986, the disclosure of which is incorporated herein by reference. Those 5′ EST sequences encoding a stretch of at least 15 amino acid long with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those 5′ ESTs which matched a known human mRNA or EST sequence and had a 5′ end more than 20 nucleotides downstream of the known 5′ end were excluded from further analysis. The remaining cDNAs having signal sequences therein were included in a database called SignalTag™.

To confirm the accuracy of the above method for identifying signal sequences, the analysis of Example 23 was performed.

EXAMPLE 23

Confirmation of Accuracy of Identification of Potential Signal Sequences in 5′ ESTs

The accuracy of the above procedure for identifying signal sequences encoding signal peptides was evaluated by applying the method to the 43 amino acids located at the N terminus of all human SwissProt proteins. The computed Von Heijne score for each protein was compared with the known characterization of the protein as being a secreted protein or a non-secreted protein. In this manner, the number of non-secreted proteins having a score higher than 3.5 (false positives) and the number of secreted proteins having a score lower than 3.5 (false negatives) could be calculated.

Using the results of the above analysis, the probability that a peptide encoded by the 5′ region of the mRNA is in fact a genuine signal peptide based on its Von Heijne's score was calculated based on either the assumption that 10% of human proteins are secreted or the assumption that 20% of human proteins are secreted. The results of this analysis are shown in Figure 2 and in table IV.

Using the above method of identification of secretory proteins, 5′ ESTs of the following polypeptides known to be secreted were obtained: human glucagon, gamma interferon induced monokine precursor, secreted cyclophilin-like protein, human pleiotropin, and human biotinidase precursor. Thus, the above method successfully identified those 5′ ESTs which encode a signal peptide.

To confirm that the signal peptide encoded by the 5′ ESTs actually functions as a signal peptide, the signal sequences from the 5′ ESTs may be cloned into a vector designed for the identification of signal peptides. Such vectors are designed to confer the ability to grow in selective medium only to host cells containing a vector with an operably linked signal sequence. For example, to confirm that a 5′ EST encodes a genuine signal peptide, the signal sequence of the 5′ EST may be inserted upstream and in frame with a non-secreted form of the yeast invertase gene in signal peptide selection vectors such as those described in U.S. Patent No. 5,536,637, the disclosure of which is incorporated herein by reference. Growth of host cells containing signal sequence selection vectors with the correctly inserted 5' EST signal sequence confirms that the 5′ EST encodes a genuine signal peptide.

Alternatively, the presence of a signal peptide may be confirmed by cloning the extended cDNAs obtained using the ESTs into expression vectors such as pXT1 (as described below in example 30), or by constructing promoter-signal sequence-reporter gene

vectors which encode fusion proteins between the signal peptide and an assayable reporter protein. After introduction of these vectors into a suitable host cell, such as COS cells or NIH 3T3 cells, the growth medium may be harvested and analyzed for the presence of the secreted protein. The medium from these cells is compared to the medium from control cells containing vectors lacking the signal sequence or extended cDNA insert to identify vectors which encode a functional signal peptide or an authentic secreted protein.

Those 5′ ESTs which encoded a signal peptide, as determined by the method of Example 22 above, were further grouped into four categories based on their homology to known sequences as described in Example 24 below.

EXAMPLE 24

Categorization of 5′ ESTs Encoding a Signal Peptide

Those 5′ ESTs having a sequence not matching any known vertebrate sequence nor any publicly available EST sequence were designated "new." Of the sequences in the SignalTag™ database, 947 of the 5′ ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5′ ESTs having a sequence not matching any vertebrate sequence but matching a publicly known EST were designated "EST-ext", provided that the known EST sequence was extended by at least 40 nucleotides in the 5′ direction. Of the sequences in the SignalTag™ database, 150 of the 5′ ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those ESTs not matching any vertebrate sequence but matching a publicly known EST without extending the known EST by at least 40 nucleotides in the 5′ direction were designated "EST." Of the sequences in the SignalTag™ database, 599 of the 5′ ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5′ ESTs matching a human mRNA sequence but extending the known sequence by at least 40 nucleotides in the 5′ direction were designated "VERT-ext." Of the sequences in the SignalTag™ database, 23 of the 5′ ESTs having a Von Heijne's score of at least 3.5 fell into this category. Included in this category was a 5′ EST which extended the known sequence of the human translocase mRNA by more than 200 bases in the 5′ direction.

A 5′ EST which extended the sequence of a human tumor suppressor gene in the 5′ direction was also identified.

Table V shows the distribution of 5′ ESTs in each category and the number of 5′ ESTs in each category having a given minimum von Heijne's score.

3. Evaluation of Spatial and Temporal Expression of mRNAs Corresponding to the 5'ESTs or Extended cDNAs

Each of the 5′ ESTs was also categorized based on the tissue from which its corresponding mRNA was obtained, as described below in Example 25.

EXAMPLE 25

Categorization of Expression Patterns

Table VI shows the distribution of 5' ESTs in each of the above defined category with respect to the tissue from which the 5'ESTs of the corresponding mRNA were obtained.

Table II provides the sequence identification numbers of 5′ EST sequences derived from prostate, the categories in which these sequences fall, and the von Heijne's score of the signal peptides which they encode. The 5′ EST sequences and the amino acid sequences they encode are provided in the appended sequence listings. Table III provides the sequence ID numbers of the 5′ ESTs and the sequences of the signal peptides which they encode. The sequences of the 5′ ESTs and the polypeptides they encode are provided in the sequence listing appended hereto.

The sequences of DNA SEQ ID NOs: 38-315 can readily be screened for any errors therein and any sequence ambiguities can be resolved by resequencing a fragment containing such errors or amibiguities on both strands. Such fragments may be obtained from the plasmids stored in the inventors' laboratory or can be isolated using the techniques described herein. Resolution of any such ambiguities or errors may be facilitated by using primers which hybridize to sequences located close to the ambiguous or erroneous sequences. For example, the primers may hybridize to sequences within 50-75 bases of the amibiguity or error. Upon resolution of an error or ambiguity, the corresponding corrections can be made in the protein sequences encoded by the DNA containing the error or amibiguity.

In addition to categorizing the 5′ ESTs with respect to their tissue of origin, the spatial and temporal expression patterns of the mRNAs corresponding to the 5′ ESTs, as well as their expression levels, may be determined as described in Example 26 below. Characterization of the spatial and temporal expression patterns and expression levels of these mRNAs is useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as will be discussed in more detail below.

Furthermore, 5′ ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from the lack of expression, over expression, or under expression of an mRNA corresponding to a 5′ EST. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disease, 5′ ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5′ ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5′ ESTs. It will also be appreciated that if desired, characterization may be delayed until extended cDNAs have been obtained rather than characterizing the ESTs themselves.

EXAMPLE 26

Evaluation of Expression Levels and Patterns of mRNAs

Corresponding to 5′ ESTs or Extended cDNAs

Expression levels and patterns of mRNAs corresponding to 5′ ESTs or extended cDNAs (obtainable as described below in example 27) may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, a 5′ EST, extended cDNA, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the 5′ EST or extended cDNA has 100 or more nucleotides. The plasmid is linearized and transcribed in the

presence of ribonucleotides comprising modified ribonucleotides (*i.e.* biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (*i.e.* RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The 5′ ESTs, extended cDNAs, or fragments thereof may also be tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which gene expression patterns must be determined. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an anchoring enzyme, having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5′ or 3′ most region of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a so-called tagging endonuclease is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short tag fragments from the cDNAs.

A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the tagging endonuclease to generate short tag fragments derived from the cDNAs in the second pool. The tags resulting from digestion of the first and second pools with the anchoring enzyme and the tagging endonuclease are ligated to one another to produce so-called ditags. In some embodiments, the ditags are concatamerized to produce ligation products containing from 2 to 200 ditags. The tag sequences are then determined and compared to the sequences of the 5′ ESTs or extended cDNAs to determine which 5′ ESTs or extended cDNAs are expressed

in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5′ ESTs or extended cDNAs in the cell, tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of full length cDNAs (*i.e.* extended cDNAs which include the coding sequence for the signal peptide, the coding sequence for the mature protein, and a stop codon), extended cDNAs, 5′ ESTs or fragments thereof of sufficient length to permit specific detection of gene expression. Preferably, the fragments are at least 15 nucleotides in length. More preferably, the fragments are at least 100 nucleotide long. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments, the fragments may be more than 500 nucleotide long.

For example, quantitative analysis of gene expression may be performed with full length cDNAs as defined below, extended cDNAs, 5′ ESTs, or fragments thereof in a complementary DNA microarray as described by Schena *et al.* (*Science* **270**:467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* **93**:10614-10619, 1996). Full length cDNAs, extended cDNAs, 5′ ESTs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with full length cDNAs, extended cDNAs, 5′ ESTs, or fragments thereof in complementary DNA arrays as described by Pietu *et al.*. (*Genome Research* 6:492-503, 1996). The full length cDNAs, extended cDNAs, 5′ ESTs or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the 5′ ESTs or extended cDNAs can be done through high density nucleotide arrays as described by Lockhart *et al.* (*Nature Biotechnology* **14**: 1675-1680, 1996) and Sosnowsky *et al.* (*Proc. Natl. Acad. Sci.* **94**:1119-1123, 1997). Oligonucleotides of 15-50 nucleotides corresponding to sequences of the 5′ ESTs or extended cDNAs are synthesized directly on the chip (Lockhart *et al.*, *supra*) or synthesized and then addressed to the chip (Sosnowsky *et al.*, *supra*). Preferably, the oligonucleotides are about 20 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al*, *supra* and application of different electric fields (Sonowsky et *al*, *supra*.), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the mRNA corresponding to the 5′ EST or extended cDNA from which the oligonucleotide sequence has been designed.

III. Use of 5′ ESTs to Clone Extended cDNAs and to Clone the Corresponding Genomic DNAs

Once 5′ ESTs which include the 5′ end of the corresponding mRNAs have been selected using the procedures described above, they can be utilized to isolate extended

cDNAs which contain sequences adjacent to the 5' ESTs. The extended cDNAs may include the entire coding sequence of the protein encoded by the corresponding mRNA, including the authentic translation start site, the signal sequence, and the sequence encoding the mature protein remaining after cleavage of the signal peptide. Such extended cDNAs are referred to herein as "full length cDNAs." Alternatively, the extended cDNAs may include only the sequence encoding the mature protein remaining after cleavage of the signal peptide, or only the sequence encoding the signal peptide.

Example 27 below describes a general method for obtaining extended cDNAs using 5' ESTs. Example 28 below provides experimental results, using the method explained in example 27, describing several extended cDNAs including the entire coding sequence and authentic 5' end of the corresponding mRNA for several secreted proteins.

The methods of Examples 27, 28, and 29 can also be used to obtain extended cDNAs which encode less than the entire coding sequence of the secreted proteins encoded by the genes corresponding to the 5' ESTs. In some embodiments, the extended cDNAs isolated using these methods encode at least 10 amino acids of one of the proteins encoded by the sequences of SEQ ID NOs: 38-315. In further embodiments, the extended cDNAs encode at least 20 amino acids of the proteins encoded by the sequences of SEQ ID NOs: 38-315. In further embodiments, the extended cDNAs encode at least 30 amino amino acids of the sequences of SEQ ID NOs: 38-315. In a preferred embodiment, the extended cDNAs encode a full length protein sequence, which includes the protein coding sequences of SEQ ID NOs: 38-315.

EXAMPLE 27

General Method for Using 5' ESTs to Clone and Sequence cDNAs which Include the Entire Coding Region and the Authentic 5' End of the Corresponding mRNA

The following general method has been used to quickly and efficiently isolate extended cDNAs having the authentic 5' ends of their corresponding mRNAs as well as the full protein coding sequence and including sequence adjacent to the sequences of the 5' ESTs used to obtain them. This method may be applied to obtain extended cDNAs for any 5' EST in the NetGene™ database, including those 5' ESTs encoding polypeptides belonging to secreted proteins. The method is summarized in figure 3.

1. Obtention of Extended cDNAs

a) First strand synthesis

The method takes advantage of the known 5′ sequence of the mRNA. A reverse transcription reaction is conducted on purified mRNA with a poly 14dT primer containing a 49 nucleotide sequence at its 5′ end allowing the addition of a known sequence at the end of the cDNA which corresponds to the 3′ end of the mRNA. For example, the primer may have the following sequence: 5′-ATC GTT GAG ACT CGT ACC AGC AGA GTC ACG AGA GAG ACT ACA CGG TAC TGG TTT TTT TTT TTT TTVN -3′ (SEQ ID NO:14). Those skilled in the art will appreciate that other sequences may also be added to the poly dT sequence and used to prime the first strand synthesis. Using this primer and a reverse transcriptase such as the Superscript II (Gibco BRL) or Rnase H Minus M-MLV (Promega) enzyme, a reverse transcript anchored at the 3′ polyA site of the RNAs is generated.

After removal of the mRNA hybridized to the first cDNA strand by alkaline hydrolysis, the products of the alkaline hydrolysis and the residual poly dT primer are eliminated with an exclusion column such as an AcA34 (Biosepra) matrix as explained in Example 11.

b) Second strand synthesis

A pair of nested primers on each end is designed based on the known 5′ sequence from the 5′ EST and the known 3′ end added by the poly dT primer used in the first strand synthesis. Softwares used to design primers are either based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, *PCR Meth. Appl.* **1**:124-128, 1991), or based on the octamer frequency disparity method (Griffais *et al.*, *Nucleic Acids Res.* **19**: 3887-3891, 1991) such as PC-Rare (http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html).

Preferably, the nested primers at the 5′ end are separated from one another by four to nine bases. The 5′ primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR.

Preferably, the nested primers at the 3′ end are separated from one another by four to nine bases. For example, the nested 3′ primers may have the following sequences: (5′- CCA GCA GAG TCA CGA GAG AGA CTA CAC GG -3′(SEQ ID NO:15), and 5′- CAC GAG AGA GAC TAC ACG GTA CTG G -3′ (SEQ ID NO:16). These primers were selected

because they have melting temperatures and specificities compatible with their use in PCR. However, those skilled in the art will appreciate that other sequences may also be used as primers.

The first PCR run of 25 cycles is performed using the Advantage Tth Polymerase Mix (Clontech) and the outer primer from each of the nested pairs. A second 20 cycle PCR using the same enzyme and the inner primer from each of the nested pairs is then performed on 1/2500 of the first PCR product. Thereafter, the primers and nucleotides are removed.

2. Sequencing of Full Length Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the translation initiation codon thus yielding a nested PCR product containing the whole coding sequence. Such a full length extended cDNA undergoes a direct cloning procedure as described in section a. However, in some cases, the second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b.

a) Nested PCR products containing complete ORFs

When the resulting nested PCR product contains the complete coding sequence, as predicted from the 5'EST sequence, it is cloned in an appropriate vector such as pED6dpc2, as described in section 3.

b) Nested PCR products containing incomplete ORFs

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain both the complete coding sequence and a PCR product containing the full coding sequence. The complete coding sequence can be assembled from several partial sequences determined directly from different PCR products as described in the following section.

Once the full coding sequence has been completely determined, new primers compatible for PCR use are designed to obtain amplicons containing the whole coding region. However, in such cases, 3' primers compatible for PCR use are located inside the

3' UTR of the corresponding mRNA, thus yielding amplicons which lack part of this region, *i.e.* the polyA tract and sometimes the polyadenylation signal, as illustrated in figure 3. Such full length extended cDNAs are then cloned into an appropriate vector as described in section 3.

c) Sequencing extended cDNAs

Sequencing of extended cDNAs is performed using a Die Terminator approach with the AmpliTaq DNA polymerase FS kit available from Perkin Elmer.

In order to sequence PCR fragments, primer walking is performed using software such as OSP to choose primers and automated computer software such as ASMG (Sutton *et al., Genome Science Technol.* 1: 9-19, 1995) to construct contigs of walking sequences including the initial 5' tag using minimum overlaps of 32 nucleotides. Preferably, primer walking is performed until the sequences of full length cDNAs are obtained.

Completion of the sequencing of a given extended cDNA fragment is assessed as follows. Since sequences located after a polyA tract are difficult to determine precisely in the case of uncloned products, sequencing and primer walking processes for PCR products are interrupted when a polyA tract is identified in extended cDNAs obtained as described in case b. The sequence length is compared to the size of the nested PCR product obtained as described above. Due to the limited accuracy of the determination of the PCR product size by gel electrophoresis, a sequence is considered complete if the size of the obtained sequence is at least 70 % the size of the first nested PCR product. If the length of the sequence determined from the computer analysis is not at least 70% of the length of the nested PCR product, these PCR products are cloned and the sequence of the insertion is determined. When Northern blot data are available, the size of the mRNA detected for a given PCR product is used to finally assess that the sequence is complete. Sequences which do not fulfill the above criteria are discarded and will undergo a new isolation procedure.

Sequence data of all extended cDNAs are then transferred to a proprietary database, where quality controls and validation steps are carried out as described in example 15.

3. Cloning of Full Length Extended cDNAs

The PCR product containing the full coding sequence is then cloned in an appropriate vector. For example, the extended cDNAs can be cloned into the expression vector pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) as follows. pED6dpc2 vector DNA is prepared with blunt ends by performing an EcoRI digestion followed by a fill in reaction. The blunt ended vector is dephosphorylated. After removal of PCR primers and ethanol precipitation, the PCR product containing the full coding sequence or the extended cDNA obtained as described above is phosphorylated with a kinase subsequently removed by phenol-Sevag extraction and precipitation. The double stranded extended cDNA is then ligated to the vector and the resulting expression plasmid introduced into appropriate host cells.

Since the PCR products obtained as described above are blunt ended molecules that can be cloned in either direction, the orientation of several clones for each PCR product is determined. Then, 4 to 10 clones are ordered in microtiter plates and subjected to a PCR reaction using a first primer located in the vector close to the cloning site and a second primer located in the portion of the extended cDNA corresponding to the 3′ end of the mRNA. This second primer may be the antisense primer used in anchored PCR in the case of direct cloning (case a) or the antisense primer located inside the 3'UTR in the case of indirect cloning (case b). Clones in which the start codon of the extended cDNA is operably linked to the promoter in the vector so as to permit expression of the protein encoded by the extended cDNA are conserved and sequenced. In addition to the ends of cDNA inserts, approximately 50 bp of vector DNA on each side of the cDNA insert are also sequenced.

The cloned PCR products are then entirely sequenced according to the aforementioned procedure. In this case, contigation of long fragments is then performed on walking sequences that have already contigated for uncloned PCR products during primer walking. Sequencing of cloned amplicons is complete when the resulting contigs include the whole coding region as well as overlapping sequences with vector DNA on both ends.

4. Computer analysis of Full Length Extended cDNA

Sequences of all full length extended cDNAs are then submitted to further analysis as described below. Before searching the extended full length cDNAs for sequences of interest, extended cDNAs which are not of interest (vector RNAs, transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs and fungal RNAs) are discarded using methods essentially similar to those described for 5'ESTs in Example 18.

a) Identification of structural features

Structural features, e.g. polyA tail and polyadenylation signal, of the sequences of full length extended cDNAs are subsequently determined as follows.

A polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it. The polyA tail search is restricted to the last 100 nt of the sequence and limited to stretches of 11 consecutive A's because sequencing reactions are often not readable after such a polyA stretch. Stretches having more than 90% homology over 8 nucleotides are identified as polyA tails using BLAST2N.

To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 bp preceding the polyA tail are first searched for the canonic polyadenylation AAUAAA signal and, if the canonic signal is not detected, for the alternative AUUAAA signal (Sheets *et al., Nuc. Acids Res.* **18**: 5799-5805, 1990). If neither of these consensus polyadenylation signals is found, the canonic motif is searched again allowing one mismatch to account for possible sequencing errors. More than 85 % of identified polyadenylation signals of either type actually ends 10 to 30 bp from the polyA tail. Alternative AUUAAA signals represents approximately 15 % of the total number of identified polyadenylation signals.

b) Identification of functional features

Functional features, e.g. ORFs and signal sequences, of the sequences of full length extended cDNAs were subsequently determined as follows.

The 3 upper strand frames of extended cDNAs are searched for ORFs defined as the maximum length fragments beginning with a translation intiation codon and ending with a stop codon. ORFs encoding at least 20 amino acids are preferred.

Each found ORF is then scanned for the presence of a signal peptide in the first 50 amino-acids or, where appropriate, within shorter regions down to 20 amino acids or

less in the ORF, using the matrix method of von Heijne (*Nuc. Acids Res.* **14**: 4683-4690, 1986), the disclosure of which is incorporated herein by reference as described in Example 22.

c) Homology to either nucleotidic or proteic sequences

Categorization of full-length sequences may be achieved using procedures essentially similar to those described for 5'ESTs in Example 24.

Extended cDNAs prepared as described above may be subsequently engineered to obtain nucleic acids which include desired portions of the extended cDNA using conventional techniques such as subcloning, PCR, or *in vitro* oligonucleotide synthesis. For example, nucleic acids which include only the full coding sequences (*i.e.* the sequences encoding the signal peptide and the mature protein remaining after the signal peptide is cleaved off) may be obtained using techniques known to those skilled in the art. Alternatively, conventional techniques may be applied to obtain nucleic acids which contain only the coding sequences for the mature protein remaining after the signal peptide is cleaved off or nucleic acids which contain only the coding sequences for the signal peptides.

Similarly, nucleic acids containing any other desired portion of the coding sequences for the secreted protein may be obtained. For example, the nucleic acid may contain at least 10 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In another embodiment, the nucleic acid may contain at least 15 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. Alternatively, the nucleic acid may contain at least 20 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In another embodiment, the nucleic acid may contain at least 25 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In yet another embodiment, the nucleic acid may contain at least 40 consecutive bases of an extended cDNA such as one of the extended cDNAs described below.

Once an extended cDNA has been obtained, it can be sequenced to determine the amino acid sequence it encodes. Once the encoded amino acid sequence has been determined, one can create and identify any of the many conceivable cDNAs that will encode that protein by simply using the degeneracy of the genetic code. For example, allelic variants

or other homologous nucleic acids can be identified as described below. Alternatively, nucleic acids encoding the desired amino acid sequence can be synthesized *in vitro*.

In a preferred embodiment, the coding sequence may be selected using the known codon or codon pair preferences for the host organism in which the cDNA is to be expressed.

The extended cDNAs derived from the 5' ESTS of the present invention were obtained as described in Example 28 below.

EXAMPLE 28

Characterization of cloned extended cDNAs obtained using 5' ESTs

The procedure described in Example 27 above was used to obtain the extended cDNAs derived from the 5' ESTs of the present invention in a variety of tissues. The following list provides a few examples of thus obtained extended cDNAs.

Using this approach, the full length cDNA of SEQ ID NO:17 (internal identification number 48-19-3-G1-FL1) was obtained. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MKKVLLLITAILAVAVG (SEQ ID NO: 18) having a von Heijne score of 8.2.

The full length cDNA of SEQ ID NO:19 (internal identification number 58-34-2-E7-FL2) was also obtained using this procedure. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MWWFQQGLSFLPSALVIWTSA (SEQ ID NO:20) having a von Heijne score of 5.5.

Another full length cDNA obtained using the procedure described above has the sequence of SEQ ID NO:21 (internal identification number 51-27-1-E8-FL1). This cDNA, falls into the "EST-ext" category described above and encodes the signal peptide MVLTTLPSANSANSPVNMPTTGPNSLSYASSALSPCLT (SEQ ID NO:22) having a von Heijne score of 5.9.

The above procedure was also used to obtain a full length cDNA having the sequence of SEQ ID NO:23 (internal identification number 76-4-1-G5-FL1). This cDNA falls into the "EST-ext" category described above and encodes the signal peptide ILSTVTALTFAXA (SEQ ID NO:24) having a von Heijne score of 5.5.

The full length cDNA of SEQ ID NO:25 (internal identification number 51-3-3-B10-FL3) was also obtained using this procedure. This cDNA falls into the "new" category

described above and encodes a signal peptide LVLTLCTLPLAVA (SEQ ID NO:26) having a von Heijne score of 10.1.

The full length cDNA of SEQ ID NO:27 (internal identification number 58-35-2-F10-FL2) was also obtained using this procedure. This cDNA falls into the "new" category described above and encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:28) having a von Heijne score of 10.7.

Bacterial clones containing plasmids containing the full length cDNAs described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the stored materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the cDNA insertion. The PCR product which corresponds to the cDNA can then be manipulated using standard cloning techniques familiar to those skilled in the art.

The polypeptides encoded by the extended cDNAs may be screened for the presence of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The conserved regions have been used to derive consensus patterns or matrices included in the PROSITE data bank, in particular in the file prosite.dat (Release 13.0 of November 1995, located at http://expasy.hcuge.ch/sprot/prosite.html. Prosite_convert and prosite_scan programs (http://ulrec3.unil.ch/ftpserveur/prosite_scan) may be used to find signatures on the extended cDNAs.

For each pattern obtained with the prosite_convert program from the prosite.dat file, the accuracy of the detection on a new protein sequence may be assessed by evaluating the frequency of irrelevant hits on the population of human secreted proteins included in the data bank SWISSPROT. The ratio between the number of hits on shuffled proteins (with a window size of 20 amino acids) and the number of hits on native (unshuffled) proteins may be

used as an index. Every pattern for which the ratio is greater than 20% (one hit on shuffled proteins for 5 hits on native proteins) may be skipped during the search with prosite_scan. The program used to shuffle protein sequences (db_shuffled) and the program used to determine the statistics for each pattern in the protein data banks (prosite_statistics) are available on the ftp site http://ulrec3.unil.ch/ftpserveur/prosite_scan.

In addition to PCR based methods for obtaining extended cDNAs, traditional hybridization based methods may also be employed. These methods may also be used to obtain the genomic DNAs which encode the mRNAs from which the 5′ ESTs were derived, mRNAs corresponding to the extended cDNAs, or nucleic acids which are homologous to extended cDNAs or 5′ ESTs. Example 29 below provides examples of such methods.

EXAMPLE 29

Methods for Obtaining cDNAs which include the Entire Coding Region and the Authentic 5'End of the Corresponding mRNA

A full length cDNA library can be made using the strategies described in Examples 13, 14, 15, and 16 above by replacing the random nonamer used in Example 14 with an oligo-dT primer. For instance, the oligonucleotide of SEQ ID NO:14 may be used.

Alternatively, a cDNA library or genomic DNA library may be obtained from a commercial source or made using techniques familiar to those skilled in the art. Such cDNA or genomic DNA librairies may be used to isolate extended cDNAs obtained from 5′ EST or nucleic acids homologous to extended cDNAs or 5′ EST as follows. The cDNA library or genomic DNA library is hybridized to a detectable probe comprising at least 10 consecutive nucleotides from the 5′ EST or extended cDNA using conventional techniques. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5′ EST or extended cDNA. More preferably, the probe comprises at least 20 to 30 consecutive nucleotides from the 5′ EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5′ EST or extended cDNA.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given probe sequence are disclosed in Sambrook *et al., Molecular Cloning: A Laboratory Manual*

2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference. The same techniques may be used to isolate genomic DNAs.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the 5′ EST or extended cDNA is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5′ EST or extended cDNA. More preferably, the probe comprises 20 to 30 consecutive nucleotides from the 5′ EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5′ EST or extended cDNA.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify extended cDNAs or genomic DNAs which hybridize to the detectable probe, extended cDNAS having different levels of homology to the probe can be identified and isolated as described below.

1. Identification of Extended cDNA or Genomic cDNA Sequences Having a High Degree of Homology to the Labeled Probe

To identify extended cDNAs or genomic DNAs having a high degree of homology to the probe sequence, the melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature (Tm) is calculated using the formula: $Tm=81.5+16.6(\log [Na+])+0.41(\text{fraction } G+C)-(600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation Tm=81.5+16.6(log [Na+])+0.41(fraction G+C)-(0.63% formamide)-(600/N) where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 μg denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al., supra.*

Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the Tm. Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Extended cDNAs, nucleic acids homologous to extended cDNAs or 5′ ESTs, or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

2. Obtention of Extended cDNA or Genomic cDNA Sequences Having Lower Degrees of Homology to the Labeled Probe

The above procedure may be modified to identify extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

Extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs which have hybridized to the probe are identified by autoradiography.

3. Determination of the Degree of Homology Between the Obtained Extended cDNAs and the Labeled Probe

If it is desired to obtain nucleic acids homologous to extended cDNAs, such as allelic variants thereof or nucleic acids encoding proteins related to the proteins encoded by the extended cDNAs, the level of homology between the hybridized nucleic acid and the extended cDNA or 5′ EST used as the probe may be further determined using BLAST2N; parameters may be adapted depending on the sequence length and degree of homology studied. To determine the level of homology between the hybridized nucleic acid and the extended cDNA or 5′EST from which the probe was derived, the nucleotide sequences of the hybridized nucleic acid and the extended cDNA or 5′EST from which the probe was derived are compared. For example, using the above methods, nucleic acids having at least 95%

nucleic acid homology to the extended cDNA or 5'EST from which the probe was derived may be obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA or 5'EST from which the probe was derived.

To determine whether a clone encodes a protein having a given amount of homology to the protein encoded by the extended cDNA or 5' EST, the amino acid sequence encoded by the extended cDNA or 5' EST is compared to the amino acid sequence encoded by the hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the extended cDNA or 5' EST is closely related to an amino acid sequence in the hybridizing nucleic acid. A sequence is closely related when it is identical to that of the extended cDNA or 5' EST or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the extended cDNA or 5'EST from which the probe was derived.

In addition to the above described methods, other protocols are available to obtain extended cDNAs using 5' ESTs as outlined in the following paragraphs.

Extended cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA strand.

The first cDNA strand is hybridized to a second primer containing at least 10 consecutive nucleotides of the sequences of SEQ ID NOs 38-315. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the sequences of SEQ ID NOs 38-315. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the sequences of SEQ ID NOs 38-315. In some embodiments, the primer comprises more than 30 nucleotides from the sequences of SEQ ID NOs 38-315. If it is desired to obtain extended

cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

Extended cDNAs containing 5′ fragments of the mRNA may be prepared by hybridizing an mRNA comprising the sequence of the 5'EST for which an extended cDNA is desired with a primer comprising at least 10 consecutive nucleotides of the sequences complementary to the 5'EST and reverse transcribing the hybridized primer to make a first cDNA strand from the mRNAs. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the 5'EST. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the 5'EST.

Thereafter, a second cDNA strand complementary to the first cDNA strand is synthesized. The second cDNA strand may be made by hybridizing a primer complementary to sequences in the first cDNA strand to the first cDNA strand and extending the primer to generate the second cDNA strand.

The double stranded extended cDNAs made using the methods described above are isolated and cloned. The extended cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described *in Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. 1997 and Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989, the entire disclosures of which are incorporated herein by reference.

Alternatively, procedures such as the one described in Example 29 may be used for obtaining full length cDNAs or extended cDNAs. In this approach, full length or extended cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by

treatment with an endonuclease, such as the Gene II product of the phage F1, and an exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993). A biotinylated oligonucleotide comprising the sequence of a 5′ EST, or a fragment containing at least 10 nucleotides thereof, is hybridized to the single stranded phagemids. Preferably, the fragment comprises at least 12, 15, or 17 consecutive nucleotides from the 5′ EST. More preferably, the fragment comprises 20-30 consecutive nucleotides from the 5′ EST. In some procedures, the fragment may comprise more than 30 consecutive nucleotides from the 5′ EST.

Hybrids between the biotinylated oligonucleotide and phagemids having inserts containing the 5′ EST sequence are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry *et al.*, *Biotechniques*, 13: 124-131, 1992). Therafter, the resulting phagemids containing the 5′ EST sequence are released from the beads and converted into double stranded DNA using a primer specific for the 5′ EST sequence. Alternatively, protocoles such as the Gene Trapper kit (Gibco BRL) may be used. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs containing the 5′ EST sequence are identified by colony PCR or colony hybridization.

Using any of the above described methods in section III, a plurality of extended cDNAs containing full length protein coding sequences or sequences encoding only the mature protein remaining after the signal peptide is cleaved off may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

IV. Expression of Proteins Encoded by Extended cDNAs Isolated Using 5′ ESTs

Extended cDNAs containing the full protein coding sequences of their corresponding mRNAs or portions thereof, such as cDNAs encoding the mature protein, may be used to express the encoded secreted proteins or portions thereof as described in Example 30 below. If desired, the extended cDNAs may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. It will be appreciated that a plurality of extended cDNAs containing the full protein coding sequences or portions thereof may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

EXAMPLE 30

Expression of the Proteins Encoded by the Genes Corresponding to 5'ESTS or Portions Thereof

To express the proteins encoded by the genes corresponding to 5' ESTs (or portions thereof), full length cDNAs containing the entire protein coding region or extended cDNAs containing sequences adjacent to the 5' ESTs (or portions thereof) are obtained as described in Examples 27-29 and cloned into a suitable expression vector. If desired, the nucleic acids may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, U.S. Patent No. 5,082,767, incorporated herein by this reference.

The cDNA cloned into the expression vector may encode the entire protein (*i.e.* the signal peptide and the mature protein), the mature protein (*i.e.* the protein created by cleaving the signal peptide off), only the signal peptide or any other portion thereof.

The following is provided as one exemplary method to express the proteins encoded by the extended cDNAs corresponding to the 5' ESTs or the nucleic acids described above. First, the methionine initiation codon for the gene and the polyA signal of the gene are identified. If the nucleic acid encoding the polypeptide to be expressed lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the extended cDNA lacks a polyA signal, this sequence can be added to the construct by, for example, splicing out the

polyA signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The extended cDNA or portion thereof encoding the polypeptide to be expressed is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the extended cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the extended cDNA is positioned with the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1 containing a poly A signal and prepared for this ligation (blunt/BglII).

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 µg/ml G418 (Sigma, St. Louis, Missouri). Preferably the expressed protein is released into the culture medium, thereby facilitating purification.

Alternatively, the extended cDNAs may be cloned into pED6dpc2 as described above. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the protein expressed from the extended cDNA is released into the culture medium thereby facilitating purification.

Proteins in the culture medium are separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to electrophoresis.

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms and the proteins in the medium are harvested. The secreted proteins present in the medium are detected using techniques familiar to those skilled in the art such as Coomassie blue or silver staining or using antibodies against the protein encoded by the extended cDNA

Antibodies capable of specifically recognizing the protein of interest may be generated using synthetic 15-mer peptides having a sequence encoded by the appropriate 5′ EST, extended cDNA, or portion thereof. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the 5′ EST, extended cDNA, or portion thereof.

Secreted proteins from the host cells or organisms containing an expression vector which contains the extended cDNA derived from a 5′ EST or a portion thereof are compared to those from the control cells or organism. The presence of a band in the medium from the cells containing the expression vector which is absent in the medium from the control cells indicates that the extended cDNA encodes a secreted protein. Generally, the band corresponding to the protein encoded by the extended cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector with an insert encoding a secreted protein or portion thereof can be compared to the proteins expressed in control host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

The protein encoded by the extended cDNA may be purified using standard immunochromatography techniques. In such procedures, a solution containing the secreted protein, such as the culture medium or a cell extract, is applied to a column having antibodies against the secreted protein attached to the chromatography matrix. The secreted protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the extended cDNA sequence or portion thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies, the coding sequence of the extended cDNA or portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be β-globin or a nickel binding polypeptide. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the β-globin gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*., (*Basic Methods in Molecular Biology*, Davis, Dibner, and Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro* Express™ Translation Kit (Stratagene).

Following expression and purification of the secreted proteins encoded by the 5′ ESTs, extended cDNAs, or fragments thereof, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 31 below. It will be appreciated that a plurality of proteins expressed from these cDNAs may be included in a panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

EXAMPLE 31

Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The proteins encoded by the 5′ ESTs, extended cDNAs, or fragments thereof are cloned into expression vectors such as those described in Example 30. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled

in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

As discussed above, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the extended cDNAs or portions thereof made according to Examples 27-29 may be evaluated to determine their physiological activities as described below.

EXAMPLE 32

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Cytokine, Cell Proliferation or Cell Differentiation Activity

As discussed above, secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein encoded by the extended cDNAs is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D,

DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the above extended cDNAs or portions thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references, which are incorporated herein by reference: *Current Protocols in Immunology*, Ed. by Coligan *et al.*., Greene Publishing Associates and Wiley-Interscience; Takai *et al. J. Immunol.* **137**:3494-3500, 1986., Bertagnolli *et al.*, J. *Immunol.* **145**:1706-1712, 1990., Bertagnolli *et al.*, *Cell. Immunol.* **133**:327-341, 1991; Bertagnolli, *et al.*, *J. Immunol.* **149**:3778-3783, 1992; Bowman *et al., J. Immunol.* **152**:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in *Current Protocols in Immunology, supra* **1**:3.12.1-3.12.14; and Schreiber In *Current Protocols in Immunology, supra* **1** : 6.8.1-6.8.8.

The proteins encoded by the cDNAs may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following references, which are incorporated herein by reference: Bottomly *et al.*, In *Current Protocols in Immunology., supra.* **1** : 6.3.1-6.3.12,; deVries *et al.*, *J. Exp. Med.* **173**:1205-1211, 1991; Moreau *et al.*, *Nature* **36**:690-692, 1988; Greenberger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**:2931-2938, 1983; Nordan, R., In *Current Protocols in Immunology., supra.* 1 : 6.6.1-6.6.5; Smith et al., *Proc. Natl. Acad. Sci. U.S.A.* **83**:1857-1861, 1986; Bennett *et al.*, in *Current Protocols in Immunology supra* 1 : 6.15.1; Ciarletta *et al.*, In *Current Protocols in Immunology. supra* **1** : 6.13.1.

The proteins encoded by the cDNAs may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in *Current Protocols in Immunology supra*; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* **137**:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988.

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 33

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs may also be evaluated for their effects as immune regulators. For example, the proteins may be evaluated for their activity to influence thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those skilled in the art including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in *Current Protocols in Immunology* , Coligan *et al.*, Eds, Greene Publishing Associates and Wiley-Interscience; Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* **78**:2488-2492, 1981; Herrmann *et al.*, *J. Immunol.* **128**:1968-1974, 1982; Handa *et al.*, *J. Immunol.* **135**:1564-1572, 1985; Takai *et al.*, *J. Immunol.* **137**:3494-3500, 1986; Takai *et al.*, *J. Immunol.* **140**:508-512, 1988; Bowman *et al.*, *J. Virology* **61**:1992-1998; Bertagnolli *et al.*, *Cell. Immunol.* **133**:327-341, 1991; Brown *et al.*, *J. Immunol.* **153**:3079-3092, 1994.

The proteins encoded by the cDNAs may also be evaluated for their effects on T-cell dependent immunoglobulin responses and isotype switching. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Maliszewski, *J. Immunol.* **144**:3028-3033, 1990; Mond *et al.* in *Current Protocols in Immunology*, **1** : 3.8.1-3.8.16, *supra*.

The proteins encoded by the cDNAs may also be evaluated for their effect on immune effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the

following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans) in *Current Protocols in Immunology, supra*; Takai *et al., J. Immunol.* **137**:3494-3500, 1986; Takai *et al., J. Immunol.* **140**:508-512, 1988; Bertagnolli *et al., J. Immunol.* **149**:3778-3783, 1992.

The proteins encoded by the cDNAs may also be evaluated for their effect on dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Guery *et al., J. Immunol.* **134**:536-544, 1995; Inaba *et al., J. Exp. Med.* **173**:549-559, 1991; Macatonia *et al., J. Immunol.* **154**:5071-5079, 1995; Porgador *et al.J. Exp. Med* **182**:255-260, 1995; Nair *et al., J. Virol.* **67**:4062-4069, 1993; Huang *et al., Science* **264**:961-965, 1994; Macatonia *et al.J. Exp. Med* **169**:1255-1264, 1989; Bhardwaj *et al., Journal of Clinical Investigation* **94**:797-807, 1994; and Inaba *et al., J. Exp. Med* **172**:631-640, 1990.

The proteins encoded by the cDNAs may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Darzynkiewicz *et al., Cytometry* **13**:795-808, 1992; Gorczyca *et al., Leukemia* **7**:659-670, 1993; Gorczyca *et al., Cancer Res.* **53**:1945-1951, 1993; Itoh *et al., Cell* **66**:233-243, 1991; Zacharchuk, *J. Immunol.* **145**:4037-4045, 1990; Zamai *et al., Cytometry* **14**:891-897, 1993; Gorczyca *et al., Int. J. Oncol.* **1**:639-648, 1992.

The proteins encoded by the cDNAs may also be evaluated for their influence on early steps of T-cell commitment and development. Numerous assays for such activity are familiar to those skilled in the art, including without limitation the assays disclosed in the following references, which are incorporated herein by references: Antica *et al., Blood* **84**:111-117, 1994; Fine *et al., Cell. Immunol.* 155:111-122, 1994; Galy *et al., Blood* **85**:2770-2778, 1995; Toki *et al., Proc. Nat. Acad Sci. USA* **88**:7548-7551, 1991.

Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency),

e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania spp., plamodium and various fungal infections such as candidiasis. Of course, in this regard, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Alternatively, proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may be used in treatment of autoimmune disorders including, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention.

Using the proteins of the invention it may also be possible to regulate immune responses either up or down.

Down regulation may involve inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active non-antigen-specific process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after the end of exposure to the tolerizing agent. Operationally, tolerance can be

demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions, such as, for example, B7 costimulation), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation, can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al., Science* 257:789-792, 1992 and Turka *et al., Proc. Natl. Acad. Sci USA*, 89:11102-11105, 1992. In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor/ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which potentially involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/pr/pr mice or NZB hybrid mice, murine autoimmuno collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *supra*, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may involve either enhancing an existing immune response or eliciting an initial immune response as shown by the following examples. For instance, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte antigens systemically.

Alternatively, antiviral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention or together with a stimulatory form of a soluble peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention and reintroducing the *in vitro* primed T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

In another application, upregulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules can be transfected with nucleic acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain and β_2 microglobulin or an MHC class II α chain and an MHC class II β chain to thereby express MHC class I or MHC class II proteins on the cell surface, respectively. Expression of the appropriate MHC class I or class II molecules in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these immune system regulator proteins or nucleic acids regulating the expression of

such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 34

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Johansson *et al. Cell. Biol.* **15**:141-151, 1995; Keller *et al., Mol. Cell. Biol.* **13**:473-486, 1993; McClanahan *et al., Blood* **81**:2903-2915, 1993.

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, Methylcellulose Colony Forming Assays, in *Culture of Hematopoietic Cells.*, Freshney, *et al.*. Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama *et al., Proc. Natl. Acad Sci. USA* **89**:5907-5911, 1992; McNiece and Briddell, in *Culture of Hematopoietic Cells, supra*; Neben *et al., Exp. Hematol.* **22**:353-359, 1994; Ploemacher and Cobblestone In *Culture of Hematopoietic Cells, supra*1-21, Spooncer *et al*, in *Culture of Hematopoietic Cells, supra*163-179 and Sutherland in *Culture of Hematopoietic Cells, supra*. 139-162.

Those proteins which exhibit hematopoiesis regulatory activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoeisis is beneficial, such as in the treatment of myeloid or lymphoid cell deficiencies. Involvement in regulating hematopoiesis is indicated even by marginal biological activity in support of colony forming cells or of factor-dependent cell lines. For example, proteins supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, indicates utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors

and/or erythroid cells. Proteins supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) may be useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression. Proteins supporting the growth and proliferation of megakaryocytes and consequently of platelets allows prevention or treatment of various platelet disorders such as thrombocytopenia, and generally may be used in place of or complementary to platelet transfusions. Proteins supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells may therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantion, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in vivo* or *ex vivo* (*i.e.*, in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy. Alternatively, as described in more detail below, genes encoding hematopoiesis regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 35

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Tissue Growth

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effect on tissue growth. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in International Patent Publication No. WO95/16035, International Patent Publication No. WO95/05846 and International Patent Publication No. WO91/07491, which are incorporated herein by reference.

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112, Maibach and Rovee, eds., Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, *J. Invest. Dermatol.* 71:382-84, 1978, which are incorporated herein by reference.

Those proteins which are involved in the regulation of tissue growth may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of tissue growth is beneficial. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. *De novo* bone synthesis induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of bone-forming cell progenitors. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein encoded by extended cDNAs derived from the 5' ESTs of the present invention is tendon/ligament formation. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. *De novo* tendon/ligament-like tissue

formation induced by a composition encoded by extended cDNAs derived from the 5' ESTs of the present invention contributes to the repair of tendon or ligaments defects of congenital, traumatic or other origin and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions encoded by extended cDNAs derived from the 5' ESTs of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.*, for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular

endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokinc damage.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding tissue growth regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 36

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Reproductive Hormones

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Vale *et al.*, *Endocrinol.* **91**:562-572, 1972; Ling *et al.*, *Nature* **321**:779-782, 1986; Vale *et al.*, *Nature* **321**:776-779, 1986; Mason *et al..*, *Nature* **318**:659-663, 1985; Forage *et al.*, *Proc. Natl. Acad. Sci. USA* **83**:3091-3095, 1986, Chapter 6.12 in *Current Protocols in Immunology*, Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Intersciece ; Taub *et al.*, *J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al.*, *APMIS* **103**:140-146, 1995; Muller *et al.*, *Eur. J. Immunol.* 25:1744-1748; Gruber *et al.*, *J. Immunol.* **152**:5860-5867, 1994; Johnston *et al.*, *J Immunol.* **153**:1762-1768, 1994.

Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in

which regulation of reproductive hormones are beneficial. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of FSH. Thus, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, genes encoding reproductive hormone regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 37

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Chemotactic/Chemokinetic Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for chemotactic/chemokinetic activity. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins

provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, Kruisbeek, Margulies, Shevach and Strober, Pub. Greene Publishing Associates and Wiley-Interscience, Chapter **6.12**: 6.12.1-6.12.28; Taub *et al., J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al., APMIS* **103**:140-146, 1995; Mueller *et al., Eur. J. Immunol.* 25:1744-1748; Gruber *et al., J. Immunol.* **152**:5860-5867, 1994; Johnston *et al. J. Immunol.*, **153**:1762-1768, 1994.

EXAMPLE 38

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Blood Clotting

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Linet *et al., J. Clin. Pharmacol.* **26**:131-140, 1986; Burdick

et al., *Thrombosis Res.* **45**:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79, 1991; Schaub, *Prostaglandins* **35**:467-474, 1988.

Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as infarction of cardiac and central nervous system vessels (e.g., stroke)). Alternatively, as described in more detail below, genes encoding blood clotting activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 39

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Involvement in Receptor/Ligand Interactions

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 7. 7.28.1-7.28.22 in *Current Protocols in Immunology*, Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA* **84**:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* **168**:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* **169**:149-160, 1989; Stoltenborg *et al.*, *J. Immunol. Methods* **175**:59-68, 1994; Stitt *et al.*, *Cell* **80**:661-670, 1995; Gyuris *et al.*, *Cell* 75:791-803, 1993.

For example, the proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include,

without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions. Alternatively, as described in more detail below, genes encoding proteins involved in receptor/ligand interactions or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 40

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Anti-Inflammatory Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions, including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome), ischemia-reperfusioninury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine- or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Alternatively, as described in more detail below, genes encoding anti-inflammatory activity proteins or nucleic

acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 41

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Tumor Inhibition Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for tumor inhibition activity. In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth. Alternatively, as described in more detail below, genes tumor inhibition activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors;

providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein. Alternatively, as described in more detail below, genes encoding proteins involved in any of the above mentioned activities or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 42

Identification of Proteins which Interact with Polypeptides Encoded by Extended cDNAs

Proteins which interact with the polypeptides encoded by cDNAs derived from the 5′ ESTs or fragments thereof, such as receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the kit which is incorporated herein by reference, the the cDNAs derived from 5′ ESTs, or fragments thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the extended cDNAs or portions thereof are inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins which interact with the polypeptide encoded by the extended cDNAs or portions thereof.

Alternatively, the system described in Lustig *et al.*, *Methods in Enzymology* **283**: 83-99, 1997, and in U.S. Patent No. 5,654,150, the disclosure of which is incorporated herein by reference, may be used for identifying molecules which interact with the polypeptides encoded by extended cDNAs. In such systems, *in vitro* transcription reactions are performed on a pool of vectors containing extended cDNA inserts cloned downstream of a promoter which drives *in vitro* transcription. The resulting pools of mRNAs are introduced into *Xenopus laevis* oocytes. The oocytes are then assayed for a desired activity.

Alternatively, the pooled *in vitro* transcription products produced as described above may be translated *in vitro*. The pooled *in vitro* translation products can be assayed for a desired activity or for interaction with a known polypeptide.

Proteins or other molecules interacting with polypeptides encoded by extended cDNAs can be found by a variety of additional techniques. In one method, affinity columns containing the polypeptide encoded by the extended cDNA or a portion thereof can be constructed. In some versions, of this method the affinity column contains chimeric proteins in which the protein encoded by the extended cDNA or a portion thereof is fused to glutathione S-transferase. A mixture of cellular proteins or pool of expressed proteins as described above and is applied to the affinity column. Proteins interacting with the polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al., Electrophoresis* **18**:588-598, 1997, the disclosure of which is incorporated herein by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

Proteins interacting with polypeptides encoded by extended cDNAs or portions thereof can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow, *Analytical Biochemistry* **246**:1-6, 1997, the disclosure of which is incorporated herein by reference. The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethl dextran matrix) and a sample of test molecules is placed in contact with

the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be one of the polypeptides encoded by extended cDNAs or a portion thereof and the test sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides. The tissues or cells from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is a collection of unique polypeptides encoded by the extended cDNAs or portions thereof.

To study the interaction of the proteins encoded by the extended cDNAs or portions thereof with drugs, the microdialysis coupled to HPLC method described by Wang *et al.*, *Chromatographia* 44:205-208, 1997 or the affinity capillary electrophoresis method described by Busch *et al.*, *J. Chromatogr.* 777:311-328, 1997, the disclosures of which are incorporated herein by reference can be used.

It will be appreciated by those skilled in the art that the proteins expressed from the extended cDNAs or portions may be assayed for numerous activities in addition to those specifically enumerated above. For example, the expressed proteins may be evaluated for applications involving control and regulation of inflammation, tumor proliferation or metastasis, infection, or other clinical conditions. In addition, the proteins expressed from the extended cDNAs or portions thereof may be useful as nutritional agents or cosmetic agents.

The proteins expressed from the cDNAs or portions thereof may be used to generate antibodies capable of specifically binding to the expressed protein or fragments thereof as described in Example 40 below. The antibodies may capable of binding a full length protein encoded by a cDNA derived from a 5′ EST, a mature protein (*i.e.* the protein generated by cleavage of the signal peptide) encoded by a cDNA derived from a 5′ EST, or a signal peptide encoded by a cDNA derived from a 5′ EST. Alternatively, the antibodies may be capable of binding fragments of at least 10 amino acids of the proteins encoded by the above cDNAs. In some embodiments, the antibodies may be capable of binding fragments of at

least 15 amino acids of the proteins encoded by the above cDNAs. In other embodiments, the antibodies may be capable of binding fragments of at least 25 amino acids of the proteins expressed from the extended cDNAs which comprise at least 25 amino acids of the proteins encoded by the above cDNAs. In further embodiments, the antibodies may be capable of binding fragments of at least 40 amino acids of the proteins encoded by the above cDNAs.

EXAMPLE 43

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 30. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few µg/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, and Milstein, *Nature* **256**:495, 1975 or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, *Meth. Enzymol.* **70**:419, 1980, the disclosure of which is incorporated herein by reference and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis *et al.* in *Basic Methods in Molecular Biology*

Elsevier, New York. Section 21-2, the disclosure of which is incorporated herein by reference.

2. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals response vary depending on site of inoculations and doses, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis. *et al, J. Clin. Endocrinol. Metab.* 33:988-991 (1971), the disclosure of which is incorporated herein by reference.

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, *et al.*, Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973), the disclosure of which is incorporated herein by reference. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980), the disclosure of which is incorporated herein by reference..

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

V. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof as Reagents

The 5′ ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may be used as reagents in isolation procedures, diagnostic assays, and forensic procedures. For example, sequences from the 5′ ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be detectably labeled and used as probes to isolate other sequences capable of hybridizing to them. In addition, sequences from 5′ ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

<u>1. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Isolation, Diagnostic and Forensic Procedures</u>

EXAMPLE 44

<u>Preparation of PCR Primers and Amplification of DNA</u>

The 5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom) may be used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. The PCR primers are at least 10 bases, and preferably at least 12, 15, or 17 bases in length. More preferably, the PCR primers are at least 20-30 bases in length. In some embodiments, the PCR primers may be more than 30 bases in length. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering, White Ed. *in Methods in Molecular Biology* **67**: Humana Press, Totowa 1997, the disclosure of which is incorporated herein by reference. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation,

hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

EXAMPLE 45

Use of 5'ESTs as Probes

Probes derived from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), including full length cDNAs or genomic sequences, may be labeled with detectable labels familiar to those skilled in the art, including radioisotopes and non-radioactive labels, to provide a detectable probe. The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described in Example 30 above.

PCR primers made as described in Example 44 above may be used in forensic analyses, such as the DNA fingerprinting techniques described in Examples 46-50 below. Such analyses may utilize detectable probes or primers based on the sequences of the the 5' ESTs or of cDNAs or genomic DNAs isolated using the 5' ESTs.

EXAMPLE 46

Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the 5′ ESTs of Example 25, or cDNAs or genomic DNAs isolated therefrom as described above, is then utilized in accordance with Example 44 to amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

EXAMPLE 47

Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of 5'EST sequences from Example 25, or cDNA or genomic DNA sequences obtainable therefrom. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 44. Each of these DNA segments is sequenced, using the methods set forth in Example 46. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

EXAMPLE 48

Southern Blot Forensic Identification

The procedure of Example 47 is repeated to obtain a panel of at least 10 amplified sequences from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis *et al.* (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65) , the disclosure of which is incorporated herein by reference..

A panel of probes based on the sequences of 5′ ESTs (or cDNAs or genomic DNAs obtainable therefrom), or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis *et al.*, supra). Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5′ EST (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5′ EST (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5′ EST (or cDNAs or genomic DNAs obtainable therefrom).

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of 5′ EST (or cDNAs or genomic DNAs obtainable therefrom) will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of 5′ EST (or cDNAs or genomic DNAs obtainable therefrom) probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 49

Dot Blot Identification Procedure

Another technique for identifying individuals using the 5′ EST sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10, preferably 50 sequences from the 5′ ESTs or cDNAs or genomic DNAs obtainable therefrom. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis et *al.*, *supra*). The ^{32}p labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA. Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood *et al.*, *Proc. Natl. Acad. Sci. USA* **82(6)**:1585-1588, 1985) which is hereby incorporated by reference. A unique pattern of dots distinguishes one individual from another individual.

5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom) or oligonucleotides containing at least 10 consecutive bases from these sequences can be used as probes in the following alternative fingerprinting technique. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom).

Preferably, a plurality of probes having sequences from different genes are used in the alternative fingerprinting technique. Example 50 below provides a representative alternative fingerprinting procedure in which the probes are derived from 5'EST.

EXAMPLE 50

Alternative "Fingerprint" Identification Technique

20-mer oligonucleotides are prepared from a large number, e.g. 50, 100, or 200, of 5'EST using commercially available oligonucleotide services such as Genset, Paris, France. Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI and XbaI. Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with ^{32}P. The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

The proteins encoded by the extended cDNAs may also be used to generate antibodies as explained in Examples 30 and 43 in order to identify the tissue type or cell species from which a sample is derived as described in example 51.

EXAMPLE 51

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Examples 30 and 43 which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, Chap. 26 in: *Basic and Clinical Immunology*, 3rd Ed. Lange, Los Altos, California, 1980, or Rose,. *et al.*, Chap. 12 in: *Methods in Immunodiagnosis*, 2d Ed. John Wiley and Sons, New York (1980), the disclosures of which are incorporated herein by reference.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I, and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm, unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative

control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of tissue specific soluble proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, *et al.*, Section 19-2 in: *Basic Methods in Molecular Biology*, Leder ed., Elsevier, New York, 1986, the disclosure of which is incorporated herein by reference, using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5

to 55 μl, and containing from about 1 to 100 μg protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. *et al.*, *supra* Section 19-3. One set of nitrocellulose blots is stained with Coomassie blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Examples 30 and 43. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from extended cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

In addition to their applications in forensics and identification, 5′ ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be mapped to their chromosomal locations. Example 52 below describes radiation hybrid (RH) mapping of human chromosomal regions using 5'ESTs. Example 53 below describes a representative procedure for mapping an 5′ EST to its location on a human chromosome. Example 54 below describes mapping of 5′ ESTs on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH). Those skilled in the art will appreciate that the method of Examples 52-54 may also be used to map cDNAs or genomic DNAs obtainable from the 5′ ESTs to their chromosomal locations.

2. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Chromosome Mapping

EXAMPLE 52

Radiation hybrid mapping of 5'ESTs to the human genome

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different portions of the human genome. This technique is described by Benham *et al.*, *Genomics* **4**:509-517, 1989; and Cox *et al.*, *Science* **250**:245-250, 1990, the entire contents of which are hereby incorporated by reference. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering 5'EST. In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler *et al.*, *Science* **274**:540-546, 1996, hereby incorporated by reference).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster *et al.*, *Genomics* **33**:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, *Eur. J. Hum. Genet.* **4**:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, *Genomics* **29**:170-178, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, *Genomics* **14**:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, *Genomics* **11**:701-708, 1991).

EXAMPLE 53

Mapping of 5'ESTs to HumanChromosomes using PCR techniques

5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the 5' ESTs (or cDNAs or genomic DNAs obtainable

therefrom) to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich in *PCR Technology; Principles and Applications for DNA Amplification*, Freeman and Co., New York, 1992, the disclosure of which is incorporated herein by reference..

The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 µCu of a ^{32}P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the extended cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given 5′ EST (or cDNA or genomic DNA obtainable therefrom). DNA is isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from the 5′ EST (or cDNA or genomic DNA obtainable therefrom). Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the 5′ EST (or cDNA or genomic DNA obtainable therefrom) will yield an amplified fragment. The 5′ EST (or cDNA or genomic DNA obtainable therefrom) are assigned to a chromosome by analysis of the segregation pattern of PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that 5′EST (or cDNA or genomic DNA obtainable therefrom). For a review of techniques and analysis of results from somatic cell gene mapping experiments, see Ledbetter *et al.*, *Genomics* 6:475-481, 1990, the disclosure of which is incorporated herein by reference.

EXAMPLE 54

Mapping of Extended 5′ ESTs to Chromosomes Using Fluorescence *In Situ* Hybridization

Fluorescence in situ hybridization allows the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence in situ hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of an 5'EST (or cDNA or genomic DNA obtainable therefrom) is obtained by FISH as described by Cherif *et al.* (*Proc. Natl. Acad. Sci. U.S.A.*, 87:6639-6643, 1990) , the disclosure of which is incorporated herein by reference.. Metaphase chromosomes are prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10 μM) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1 μg/ml) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The 5'EST (or cDNA or genomic DNA obtainable therefrom) is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upsala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70°C for 5-10 min.

Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100 μg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10 μg/100 ml in 20 mM Tris-HCl, 2 mM $CaCl_2$) at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin

and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, *supra.*). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) may be localized to a particular cytogenetic R-band on a given chromosome.

Once the 5'EST (or cDNA or genomic DNA obtainable therefrom) have been assigned to particular chromosomes using the techniques described in Examples 52-54 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

EXAMPLE 55

Use of 5'EST to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the extended cDNAs (or genomic DNAs obtainable therefrom) are obtained. This approach is described in Nagaraja *et al., Genome Research* 7:210-222, 1997, the disclosure of which is incorporated herein by reference. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector. The YAC inserts are screened using PCR or other methods to determine whether they include the 5'EST (or cDNA or genomic DNA obtainable therefrom) whose position is to be determined. Once an insert has been found which includes the 5'EST (or cDNA or genomic DNA obtainable therefrom), the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome or in the region from which the 5'EST (or cDNA or genomic DNA obtainable therefrom) was derived. This process can be repeated for each insert in the YAC library to determine the

location of each of the extended cDNAs (or genomic DNAs obtainable therefrom) relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

As described in Example 56 below extended cDNAs (or genomic DNAs obtainable therefrom) may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

3. Use of 5'ESTs or Sequences Obtained Therefrom or Fragments Thereof in Gene Identification

EXAMPLE 56

Identification of genes associated with hereditary diseases or drug response

This example illustrates an approach useful for the association of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) with particular phenotypic characteristics. In this example, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) is used as a test probe to associate that 5'EST (or cDNA or genomic DNA obtainable therefrom) with a particular phenotypic characteristic.

5'ESTs (or cDNA or genomic DNA obtainable therefrom) are mapped to a particular location on a human chromosome using techniques such as those described in Examples 52 and 53 or other techniques known in the art. A search of Mendelian Inheritance in Man (McKusick in *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which contains the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes have not been identified. The gene corresponding to this 5'EST (or cDNA or genomic DNA obtainable therefrom) thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the 5'EST (or cDNA or genomic DNA obtainable therefrom) are used to screen genomic DNA, mRNA or cDNA obtained from the

patients. 5'ESTs (or cDNA or genomic DNA obtainable therefrom) that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when the samples are derived from an individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the 5'EST may be responsible for the genetic disease.

VI. Use of 5'EST (or cDNA or Genomic DNA Obtainable Therefrom) to Construct Vectors

The present 5'ESTs (or cDNA or genomic DNA obtainable therefrom) may also be used to construct secretion vectors capable of directing the secretion of the proteins encoded by genes therein. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described in Example 57 below.

1. Construction of Secretion Vectors

EXAMPLE 57

Construction of Secretion Vectors

The secretion vectors include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

A signal sequence from a 5′ EST (or cDNAs or genomic DNAs obtainable therefrom) is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the 5' EST (or cDNA or genomic DNA obtainable therefrom). Suitable hosts include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and HPLC. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

The 5′ ESTs may also be used to clone sequences located upstream of the 5′ ESTs which are capable of regulating gene expression, including promoter sequences, enhancer sequences, and other upstream sequences which influence transcription or translation levels. Once identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion. Example 58 describes a method for cloning sequences upstream of the extended cDNAs or 5′ ESTs.

2. Identification of Upstream Sequences With Promoting or Regulatory Activities

EXAMPLE 58

Use of Extended cDNAs or 5′ ESTs to Clone Upstream Sequences from Genomic DNA

Sequences derived from extended cDNAs or 5′ ESTs may be used to isolate the promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the GenomeWalker™ kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions (which are incorporated herein by reference) using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for the extended cDNA or 5′ EST of interest and should have a melting temperature, length, and location in the extended cDNA or 5'EST which is consistent with its use in PCR reactions. Each first PCR reaction contains 5 ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adaptor primer and outer gene specific primer, 1.1 mM of $Mg(OAc)_2$, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min - 94°C / 2 sec - 94°C, 3 min - 72°C (7 cycles) / 2 sec - 94°C, 3 min - 67°C (32 cycles) / 5 min - 67°C.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested

primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first nested primer is specific for the adaptor, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular extended cDNA or 5′ EST for which the promoter is to be cloned and should have a melting temperature, length, and location in the extended cDNA or 5′ EST which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min - 94°C / 2 sec - 94°C, 3 min - 72°C (6 cycles) / 2 sec - 94°C, 3 min - 67°C (25 cycles) / 5 min - 67°C. The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques.

Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the extended cDNA or 5′ EST sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the extended cDNA or EST sequence are isolated as described in Example 29 above. Thereafter, the single stranded DNA containing the extended cDNA or EST sequence is released from the beads and converted into double stranded DNA using a primer specific for the extended cDNA or 5′ EST sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the 5′ EST or extended cDNA sequences are identified by colony PCR or colony hybridization.

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the extended cDNAs or 5′ ESTs with databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example .

EXAMPLE 59

Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the extended cDNAs or 5′ ESTs are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the extended cDNAs or 5′ ESTs are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Appropriate host cells for the promoter reporter vectors may be chosen based on the results of the above described determination of expression patterns of the extended cDNAs and ESTs. For example, if the expression pattern analysis indicates that the mRNA corresponding to a particular extended cDNA or 5′ EST is expressed in fibroblasts, the promoter reporter vector may be introduced into a human fibroblast cell line.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed

mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

EXAMPLE 60

Cloning and Identification of Promoters

Using the method described in Example 58 above with 5′ ESTs, sequences upstream of several genes were obtained. Using the primer pairs GGG AAG ATG GAG ATA GTA TTG CCT G (SEQ ID NO:29) and CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ ID NO:30), the promoter having the internal designation P13H2 (SEQ ID NO:31) was obtained.

Using the primer pairs GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO:32) and CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO:33), the promoter having the internal designation P15B4 (SEQ ID NO:34) was obtained.

Using the primer pairs CTG GGA TGG AAG GCA CGG TA (SEQ ID NO:35) and GAG ACC ACA CAG CTA GAC AA (SEQ ID NO:36), the promoter having the internal designation P29B6 (SEQ ID NO:37) was obtained.

Figure 4 provides a schematic description of the promoters isolated and the way they are assembled with the corresponding 5′ tags. The upstream sequences were screened for the presence of motifs resembling transcription factor binding sites or known transcription start sites using the computer program MatInspector release 2.0, August 1996.

Table VII describes the transcription factor binding sites present in each of these promoters. The columns labeled matrice provides the name of the MatInspector matrix used. The column labeled position provides the 5′ position of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5′ EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5′ EST. The column labeled "score" provides the MatInspector score found for this site. The column labeled "length" provides the length

of the site in nucleotides. The column labeled "sequence" provides the sequence of the site found.

Bacterial clones containing plasmids containing the promoter sequences described above described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the deposited materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5′ EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

The promoters and other regulatory sequences located upstream of the extended cDNAs or 5′ ESTs may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and quantitative patterns may be selected using the results of the expression analysis described in Example 26 above. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of an extended cDNA or 5′ EST derived from an mRNA which is expressed at a high level in muscle, as determined by the method of Example 26, may be used in the expression vector.

Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Following the identification of promoter sequences using the procedures of Examples 58-60, proteins which interact with the promoter may be identified as described in Example 61 below.

EXAMPLE 61

Identification of Proteins Which Interact with Promoter Sequences, Upstream Regulatory Sequences, or mRNA

Sequences within the promoter region which are likely to bind transcription factors may be identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter sequence. For example, deletions may be made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed. Transcription factor binding sites within the regions in which deletions reduce expression levels may be further localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the art.

Nucleic acids encoding proteins which interact with sequences in the promoter may be identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog No. K1603-1), the disclosure of which is incorporated herein by reference. Briefly, the Matchmaker One-hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GAL4, is transformed into the yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to

select cells expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts may be inserted into expression vectors or *in vitro* transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNAse protection analysis.

VII. Use of 5′ ESTs (or cDNAs or Genomic DNAs Obtainable Therefrom) in Gene Therapy

The present invention also comprises the use of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) in gene therapy strategies, including antisense and triple helix strategies as described in Examples 62 and 63 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

EXAMPLE 62

Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom). The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex with sufficient stability to inhibit the expression of the mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et*

al., *Ann. Rev. Biochem.* 55:569-597, 1986; and Izant and Weintraub, *Cell* **36**:1007-1015, 1984, which are hereby incorporated by reference.

In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of the antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity. Examples of modifications suitable for use in antisense strategies are described by Rossi *et al.*, *Pharmacol. Ther.* **50(2)**:245-254, 1991, which is hereby incorporated by reference.

Various types of antisense oligonucleotides complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom) may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3′ end or both the 3′ and 5′ ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages,

wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefore. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection, transfection or h-region-mediated import using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors,

vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between $1x10^{-10}M$ to $1x10^{-4}M$. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use *in vivo*. For example, an inhibiting concentration in culture of $1x10^{-7}$ translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi *et al.*, *supra*.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

The 5′ ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The 5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom) of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of 5′ EST sequences (or cDNAs or genomic DNAs obtainable therefrom) can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at

homopurine:homopyrimidine sequences. Thus, both types of sequences from the 5'EST or from the gene corresponding to the 5'EST are contemplated within the scope of this invention.

EXAMPLE 63

Preparation and Use of Triple Helix Probes

The sequences of the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) are scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting gene expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting gene expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which normally express the target gene. The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis, such as GENSET, Paris, France.

The oligonucleotides may be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced gene expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the target gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the extended cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the extended cDNA is associated with the disease using techniques described in Example 56.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques described above and in Example 62 at a dosage calculated based on the *in vitro* results, as described in Example 62.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3′ end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.*, *Science* 245:967-971, 1989, which is hereby incorporated by this reference.

EXAMPLE 64

Use of cDNAs Obtained Using the 5′ ESTs to Express an Encoded Protein in a Host Organism

The cDNAs obtained as described above using the 5′ ESTs of the present invention may also be used to express an encoded protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded protein may have any of the activities described above. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

A full length extended cDNA encoding the signal peptide and the mature protein, or an extended cDNA encoding only the mature protein is introduced into the host organism. The extended cDNA may be introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the extended cDNA may be injected into the host organism as naked DNA such that the encoded protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the extended cDNA may be cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector may be any of the expression vectors designed for use in gene therapy, including viral or retroviral vectors. The expression vector may be directly introduced into the host organism such that the encoded protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector may be introduced into cells *in vitro*. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the encoded protein to produce a beneficial effect.

EXAMPLE 65

Use of Signal Peptides Encoded by 5' ESTs or Sequences obtained Therefrom to Import Proteins Into Cells

The short core hydrophobic region (h) of signal peptides encoded by the 5'ESTS or extended cDNAs derived from SEQ ID NOs: 38-315 may also be used as a carrier to import a peptide or a protein of interest, so-called cargo, into tissue culture cells (Lin *et al., J. Biol. Chem.*, **270**: 14225-14258, 1995; Du *et al., J. Peptide Res.*, **51**: 235-243, 1998; Rojas *et al., Nature Biotech.*, **16**: 370-375, 1998).

When cell permeable peptides of limited size (approximately up to 25 amino acids) are to be translocated across cell membrane, chemical synthesis may be used in order to add the h region to either the C-terminus or the N-terminus to the cargo peptide of interest. Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can be genetically engineered, using techniques familiar to those skilled in the art, in order to link the extended cDNA sequence encoding the h region to the 5' or the 3' end of a DNA sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then simply incubated with the cell permeable polypeptide which is then translocated across the membrane.

This method may be applied to study diverse intracellular functions and cellular processes. For instance, it has been used to probe functionally relevant domains of intracellular proteins and to examine protein-protein interactions involved in signal transduction pathways (Lin *et al., supra*; Lin *et al., J. Biol. Chem.*, **271**: 5305-5308, 1996; Rojas *et al., J. Biol. Chem.*, **271**: 27456-27461, 1996; Liu *et al., Proc. Natl. Acad. Sci. USA*, **93**: 11819-11824, 1996; Rojas *et al., Bioch. Biophys. Res. Commun.*, **234**: 675-680, 1997).

Such techniques may be used in cellular therapy to import proteins producing therapeutic effects. For instance, cells isolated from a patient may be treated with imported therapeutic proteins and then re-introduced into the host organism.

Alternatively, the h region of signal peptides of the present invention could be used in combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form

triple helixes, as described in examples 62 and 63 respectively, in order to inhibit processing and/or maturation of a target cellular RNA.

As discussed above, the cDNAs or portions thereof obtained using the 5′ ESTs of the present invention can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, *Cell* 75:791-803, 1993, the disclosure of which is hereby incorporated by reference) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins

involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation *Molecular Cloning; A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Sambrook,. Fritsch and Maniatis eds., 1989, and *Methods in Enzymology; Guide to Molecular Cloning Techniques*, Academic Press, Berger and Kimmel eds., 1987.

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims. All documents cited herein are incorporated herein by reference in their entirety.

	Search characteristic		Selection Characteristics		
Step	Program	Strand	Parameters	Identity (%)	Length (bp)
miscellanaeous	blastn	both	S=61 X=16	90	17
tRNA	fasta	both	-	80	60
rRNA	blastn	both	S=108	80	40
mtRNA	blastn	both	S=108	80	40
Procaryotic	blastn	both	S=144	90	40
Fungal	blastn	both	S=144	90	40
Alu	fasta*	both	-	70	40
L1	blastn	both	S=72	70	40
Repeats	blastn	both	S=72	70	40
Promoters	blastn	top	S=54 X=16	90	15†
Vertebrate	fasta*	both	S=108	90	30
ESTs	blastn	both	S=108 X=16	90	30
Proteins	blastx¤	top	E = 0.001	-	-

Table 1: Parameters used for each step of EST analysis

* use "Quick Fast" Database scanner

† alignement further constrained to begin closer than 10bp to EST\5' end

¤ using BLOSUM62 substitution matrix

TABLE II

SEQ. ID NO.	CATEGORY	VON HEIJNE SCORE	TISSUE SOURCE	INTERNAL DESIGNATION
ID38	new	11.4	Cancerous prostate	76-36-2-G4-PU
ID39	new	11.3	Normal prostate	78-26-1-A7-PU
ID40	new	11	Normal prostate	78-4-3-G8-PU
ID41	new	10.7	Hypertrophic prostate	77-16-3-D7-PU
ID42	new	10.7	Hypertrophic prostate	77-7-1-H9-PU
ID43	new	10.6	Hypertrophic prostate	77-42-1-D10-PU
ID44	new	10.6	Cancerous prostate	76-34-4-C6-PU
ID45	new	10.4	Normal prostate	78-31-3-B8-PU
ID46	new	10.2	Normal prostate	78-38-1-C10-PU
ID47	new	10.2	Cancerous prostate	76-16-4-D5-PU
ID48	new	9	Hypertrophic prostate	77-38-2-B9-PU
ID49	new	8.8	Normal prostate	78-30-1-G12-PU
ID50	new	8.6	Prostate	60-17-1-F1-PU
ID51	new	8.5	Prostate	60-17-3-G8-PU
ID52	new	8.3	Normal prostate	78-8-2-H8-PU
ID53	new	8.3	Normal prostate	78-26-2-A1-PU
ID54	new	8.3	Cancerous prostate	76-23-2-B10-PU
ID55	new	8.2	Cancerous prostate	76-23-4-H9-PU
ID56	new	8.1	Normal prostate	78-44-2-C3-PU
ID57	new	8	Hypertrophic prostate	77-37-1-H3-PU
ID58	new	8	Normal prostate	78-35-2-G12-PU
ID59	new	7.8	Normal prostate	78-17-4-G2-PU
ID60	new	7.7	Normal prostate	78-5-4-F7-PU
ID61	new	7.6	Normal prostate	78-16-3-E2-PU
ID62	new	7.6	Hypertrophic prostate	77-5-1-B6-PU
ID63	new	7.6	Normal prostate	78-26-1-B5-PU
ID64	new	7.5	Cancerous prostate	76-12-1-B1-PU
ID65	new	7.5	Normal prostate	78-4-4-E7-PU
ID66	new	7.2	Hypertrophic prostate	77-11-1-A3-PU
ID67	new	7.2	Hypertrophic prostate	77-5-4-G9-PU
ID68	new	7.2	Normal prostate	78-23-4-H11-PU
ID69	new	7.2	Hypertrophic prostate	77-39-3-H7-PU
ID70	new	7.2	Cancerous prostate	76-23-4-H2-PU
ID71	new	7.2	Cancerous prostate	76-24-1-F8-PU
ID72	new	7	Normal prostate	78-39-4-D2-PU
ID73	new	7	Normal prostate	78-28-3-D2-PU
ID74	new	7	Normal prostate	78-29-3-H11-PU
ID75	new	7	Normal prostate	78-40-3-G2-PU
ID76	new	7	Cancerous prostate	76-1-2-F8-PU
ID77	new	7	Normal prostate	78-13-4-B10-PU
ID78	new	6.9	Cancerous prostate	76-12-1-A9-PU
ID79	new	6.9	Normal prostate	78-20-3-C11-PU
ID80	new	6.9	Cancerous prostate	76-9-2-D10-PU
ID81	new	6.8	Normal prostate	78-6-2-D12-PU
ID82	new	6.7	Hypertrophic prostate	77-10-1-C8-PU
ID83	new	6.7	Cancerous prostate	76-13-2-F11-PU
ID84	new	6.7	Cancerous prostate	76-4-1-G5-PU
ID85	new	6.5	Normal prostate	78-3-4-B8-PU
ID86	new	6.4	Prostate	60-11-3-G2-PU
ID87	new	6.3	Normal prostate	78-25-1-G5-PU
ID88	new	6.3	Normal prostate	78-2-2-G5-PU

SEQ. ID NO.	CATEGORY	VON HEIJNE SCORE	TISSUE SOURCE	INTERNAL DESIGNATION
ID89	new	6.3	Cancerous prostate	76-7-3-A1-PU
ID90	new	6.3	Hypertrophic prostate	77-5-1-C2-PU
ID91	new	6.2	Normal prostate	78-49-2-A11-PU
ID92	new	6.1	Normal prostate	78-7-1-B9-PU
ID93	new	6	Normal prostate	78-39-4-G3-PU
ID94	new	6	Normal prostate	78-32-2-H6-PU
ID95	new	5.9	Cancerous prostate	76-30-3-H2-PU
ID96	new	5.9	Normal prostate	78-24-3-H4-PU
ID97	new	5.9	Cancerous prostate	76-43-3-B6-PU
ID98	new	5.8	Prostate	60-16-3-A3-PU
ID99	new	5.8	Cancerous prostate	76-20-4-C11-PU
ID100	new	5.7	Cancerous prostate	76-11-1-C5-PU
ID101	new	5.7	Hypertrophic prostate	77-37-3-C1-PU
ID102	new	5.7	Prostate	60-13-2-B5-PU
ID103	new	5.7	Normal prostate	78-49-4-E4-PU
ID104	new	5.6	Normal prostate	78-37-4-C11-PU
ID105	new	5.6	Prostate	60-17-1-D8-PU
ID106	new	5.5	Normal prostate	78-36-3-D7-PU
ID107	new	5.5	Cancerous prostate	76-24-3-E11-PU
ID108	new	5.5	Prostate	60-14-2-A7-PU
ID109	new	5.4	Hypertrophic prostate	77-10-4-F9-PU
ID110	new	5.3	Cancerous prostate	76-23-3-G5-PU
ID111	new	5.3	Normal prostate	78-42-3-D3-PU
ID112	new	5.3	Prostate	60-12-1-H1-PU
ID113	new	5.3	Hypertrophic prostate	77-5-2-A3-PU
ID114	new	5.2	Normal prostate	78-37-2-G12-PU
ID115	new	5.2	Cancerous prostate	76-39-2-H1-PU
ID116	new	5.1	Prostate	60-12-3-C2-PU
ID117	new	5.1	Normal prostate	78-25-1-F11-PU
ID118	new	5.1	Normal prostate	78-36-2-C10-PU
ID119	new	5.1	Hypertrophic prostate	77-13-1-B7-PU
ID120	new	5.1	Hypertrophic prostate	77-4-4-H7-PU
ID121	new	5	Normal prostate	78-33-4-F9-PU
ID122	new	5	Cancerous prostate	76-21-1-D5-PU
ID123	new	4.8	Normal prostate	78-3-4-B3-PU
ID124	new	4.8	Cancerous prostate	76-29-4-B3-PU
ID125	new	4.8	Normal prostate	78-46-3-C6-PU
ID126	new	4.8	Hypertrophic prostate	77-13-3-F8-PU
ID127	new	4.7	Cancerous prostate	76-12-4-C3-PU
ID128	new	4.7	Cancerous prostate	76-34-4-C1-PU
ID129	new	4.7	Normal prostate	78-42-4-D2-PU
ID130	new	4.7	Cancerous prostate	76-38-2-H9-PU
ID131	new	4.6	Normal prostate	78-49-4-B5-PU
ID132	new	4.6	Cancerous prostate	76-1-1-E3-PU
ID133	new	4.6	Normal prostate	78-46-3-C4-PU
ID134	new	4.5	Cancerous prostate	76-22-2-D2-PU
ID135	new	4.5	Prostate	60-11-4-F6-PU
ID136	new	4.5	Normal prostate	78-32-2-G1-PU
ID137	new	4.4	Prostate	60-14-3-C7-PU
ID138	new	4.4	Hypertrophic prostate	77-3-4-H3-PU
ID139	new	4.4	Normal prostate	78-36-4-E12-PU
ID140	new	4.3	Hypertrophic prostate	77-42-1-A9-PU
ID141	new	4.3	Normal prostate	78-23-2-H3-PU

SEQ. ID NO.	CATEGORY	VON HEIJNE SCORE	TISSUE SOURCE	INTERNAL DESIGNATION
ID142	new	4.2	Cancerous prostate	76-39-3-C11-PU
ID143	new	4.2	Normal prostate	78-23-3-D10-PU
ID144	new	4.2	Cancerous prostate	76-32-2-B7-PU
ID145	new	4.2	Normal prostate	78-40-1-G9-PU
ID146	new	4.2	Prostate	60-12-1-E11-PU
ID147	new	4.1	Cancerous prostate	76-27-3-A6-PU
ID148	new	4	Cancerous prostate	76-43-3-B2-PU
ID149	new	4	Normal prostate	78-18-3-B4-PU
ID150	new	4	Normal prostate	78-41-2-D11-PU
ID151	new	4	Normal prostate	78-34-2-G9-PU
ID152	new	4	Normal prostate	78-4-3-G2-PU
ID153	new	4	Hypertrophic prostate	77-22-2-G2-PU
ID154	new	3.9	Cancerous prostate	76-4-4-F6-PU
ID155	new	3.9	Hypertrophic prostate	77-40-3-E10-PU
ID156	new	3.9	Normal prostate	78-10-1-H5-PU
ID157	new	3.9	Normal prostate	78-6-2-E3-PU
ID158	new	3.9	Hypertrophic prostate	77-20-3-E5-PU
ID159	new	3.9	Normal prostate	78-38-2-B5-PU
ID160	new	3.8	Prostate	60-11-2-G12-PU
ID161	new	3.8	Cancerous prostate	76-44-3-E8-PU
ID162	new	3.8	Normal prostate	78-41-3-A2-PU
ID163	new	3.7	Cancerous prostate	76-20-4-E7-PU
ID164	new	3.7	Cancerous prostate	76-17-1-E4-PU
ID165	new	3.7	Normal prostate	78-5-2-D2-PU
ID166	new	3.7	Prostate	60-11-3-B11-PU
ID167	new	3.7	Hypertrophic prostate	77-21-2-F1-PU
ID168	new	3.6	Prostate	60-12-1-A5-PU
ID169	new	3.6	Cancerous prostate	76-18-2-G12-PU
ID170	new	3.6	Normal prostate	78-7-1-G5-PU
ID171	new	3.6	Cancerous prostate	76-37-4-A5-PU
ID172	new	3.5	Normal prostate	78-50-4-A2-PU
ID173	new	3.5	Normal prostate	78-43-2-H10-PU
ID174	new	3.5	Normal prostate	78-44-3-B6-PU
ID175	new	3.5	Cancerous prostate	76-10-1-D6-PU
ID176	new	3.5	Prostate	60-11-4-F2-PU
ID177	new	3.5	Cancerous prostate	76-45-2-B12-PU
ID178	ext-est-not-vrt	14.8	Normal prostate	78-34-3-D9-PU
ID179	ext-est-not-vrt	13.6	Normal prostate	78-46-4-F4-PU
ID180	ext-est-not-vrt	12.7	Normal prostate	78-8-3-D9-PU
ID181	ext-est-not-vrt	8.8	Prostate	60-15-4-F6-PU
ID182	ext-est-not-vrt	8.5	Normal prostate	78-8-3-E6-PU
ID183	ext-est-not-vrt	7.3	Normal prostate	78-7-3-A4-PU
ID184	ext-est-not-vrt	7.1	Cancerous prostate	76-33-2-F5-PU
ID185	ext-est-not-vrt	6.6	Cancerous prostate	76-34-4-G12-PU
ID186	ext-est-not-vrt	6.3	Normal prostate	78-13-1-H7-PU
ID187	ext-est-not-vrt	5.9	Normal prostate	78-49-3-B11-PU
ID188	ext-est-not-vrt	5.9	Normal prostate	78-42-2-A10-PU
ID189	ext-est-not-vrt	5.5	Cancerous prostate	76-7-4-D9-PU
ID190	ext-est-not-vrt	5.2	Normal prostate	78-40-3-B12-PU
ID191	ext-est-not-vrt	5	Hypertrophic prostate	77-36-1-G2-PU
ID192	ext-est-not-vrt	4.8	Prostate	60-17-3-H11-PU
ID193	ext-est-not-vrt	4.4	Normal prostate	78-28-3-E4-PU
ID194	ext-est-not-vrt	4.1	Cancerous prostate	76-28-2-H5-PU

SEQ. ID NO.	CATEGORY	VON HEIJNE SCORE	TISSUE SOURCE	INTERNAL DESIGNATION
ID195	ext-est-not-vrt	4.1	Normal prostate	78-27-1-D11-PU
ID196	ext-est-not-vrt	3.9	Cancerous prostate	76-42-2-B5-PU
ID197	ext-est-not-vrt	3.9	Hypertrophic prostate	77-39-3-F8-PU
ID198	ext-est-not-vrt	3.7	Cancerous prostate	76-43-1-G9-PU
ID199	est-not-ext	13.8	Normal prostate	78-40-1-B10-PU
ID200	est-not-ext	13.4	Cancerous prostate	76-15-1-F4-PU
ID201	est-not-ext	13	Cancerous prostate	76-45-4-E7-PU
ID202	est-not-ext	11.6	Normal prostate	78-26-2-H7-PU
ID203	est-not-ext	11.2	Normal prostate	78-21-1-B7-PU
ID204	est-not-ext	11.2	Cancerous prostate	76-40-2-F5-PU
ID205	est-not-ext	10.6	Cancerous prostate	76-29-2-G8-PU
ID206	est-not-ext	10.5	Hypertrophic prostate	77-23-4-H11-PU
ID207	est-not-ext	10.3	Normal prostate	78-48-1-F10-PU
ID208	est-not-ext	9.5	Cancerous prostate	76-41-4-G9-PU
ID209	est-not-ext	9.3	Hypertrophic prostate	77-3-3-C10-PU
ID210	est-not-ext	9.1	Cancerous prostate	76-45-4-C8-PU
ID211	est-not-ext	8.8	Normal prostate	78-50-4-C10-PU
ID212	est-not-ext	8.8	Normal prostate	78-38-4-F7-PU
ID213	est-not-ext	8.6	Cancerous prostate	76-16-4-C9-PU
ID214	est-not-ext	8.6	Normal prostate	78-49-2-D10-PU
ID215	est-not-ext	8.4	Cancerous prostate	76-1-1-H7-PU
ID216	est-not-ext	7.9	Normal prostate	78-4-2-F10-PU
ID217	est-not-ext	7.9	Normal prostate	78-46-3-B6-PU
ID218	est-not-ext	7.7	Normal prostate	78-7-1-F2-PU
ID219	est-not-ext	7.6	Normal prostate	78-35-2-D3-PU
ID220	est-not-ext	7.6	Cancerous prostate	76-20-2-G7-PU
ID221	est-not-ext	7.6	Normal prostate	78-39-1-E11-PU
ID222	est-not-ext	7.5	Cancerous prostate	76-4-4-C2-PU
ID223	est-not-ext	7.1	Normal prostate	78-48-2-F6-PU
ID224	est-not-ext	7	Cancerous prostate	76-32-4-A10-PU
ID225	est-not-ext	6.8	Cancerous prostate	76-39-1-E7-PU
ID226	est-not-ext	6.7	Cancerous prostate	76-29-4-E1-PU
ID227	est-not-ext	6.7	Normal prostate	78-28-4-B9-PU
ID228	est-not-ext	6.7	Normal prostate	78-37-4-B2-PU
ID229	est-not-ext	6.7	Normal prostate	78-50-2-E12-PU
ID230	est-not-ext	6.7	Hypertrophic prostate	77-21-2-F8-PU
ID231	est-not-ext	6.6	Normal prostate	78-27-4-E2-PU
ID232	est-not-ext	6.5	Normal prostate	78-45-4-G12-PU
ID233	est-not-ext	6.3	Cancerous prostate	76-7-4-H8-PU
ID234	est-not-ext	6.3	Normal prostate	78-23-1-D10-PU
ID235	est-not-ext	6.3	Cancerous prostate	76-34-1-C2-PU
ID236	est-not-ext	6.2	Hypertrophic prostate	77-8-1-F11-PU
ID237	est-not-ext	6.2	Cancerous prostate	76-41-1-F3-PU
ID238	est-not-ext	6.1	Cancerous prostate	76-22-3-G4-PU
ID239	est-not-ext	6.1	Normal prostate	78-40-1-A6-PU
ID240	est-not-ext	6	Normal prostate	78-41-2-H11-PU
ID241	est-not-ext	6	Normal prostate	78-6-3-A12-PU
ID242	est-not-ext	6	Hypertrophic prostate	77-25-1-A6-PU
ID243	est-not-ext	5.9	Hypertrophic prostate	77-35-2-E4-PU
ID244	est-not-ext	5.9	Hypertrophic prostate	77-36-1-G4-PU
ID245	est-not-ext	5.8	Hypertrophic prostate	77-40-3-D6-PU
ID246	est-not-ext	5.8	Normal prostate	78-17-3-A3-PU
ID247	est-not-ext	5.7	Normal prostate	78-33-3-D7-PU

SEQ. ID NO.	CATEGORY	VON HEIJNE SCORE	TISSUE SOURCE	INTERNAL DESIGNATION
ID248	est-not-ext	5.7	Hypertrophic prostate	77-23-4-E10-PU
ID249	est-not-ext	5.7	Cancerous prostate	76-25-4-F11-PU
ID250	est-not-ext	5.7	Cancerous prostate	76-33-2-F8-PU
ID251	est-not-ext	5.7	Normal prostate	78-47-4-D6-PU
ID252	est-not-ext	5.7	Normal prostate	78-34-4-G6-PU
ID253	est-not-ext	5.6	Cancerous prostate	76-23-3-G8-PU
ID254	est-not-ext	5.6	Normal prostate	78-41-1-A6-PU
ID255	est-not-ext	5.6	Cancerous prostate	76-38-1-E4-PU
ID256	est-not-ext	5.5	Normal prostate	78-2-4-F11-PU
ID257	est-not-ext	5.4	Cancerous prostate	76-13-3-A9-PU
ID258	est-not-ext	5.4	Normal prostate	78-7-3-D9-PU
ID259	est-not-ext	5.2	Cancerous prostate	76-6-2-G5-PU
ID260	est-not-ext	5.1	Hypertrophic prostate	77-39-4-H4-PU
ID261	est-not-ext	5	Hypertrophic prostate	77-13-3-F1-PU
ID262	est-not-ext	5	Normal prostate	78-24-4-A4-PU
ID263	est-not-ext	4.9	Hypertrophic prostate	77-1-2-B4-PU
ID264	est-not-ext	4.9	Cancerous prostate	76-42-2-F3-PU
ID265	est-not-ext	4.9	Cancerous prostate	76-40-3-G6-PU
ID266	est-not-ext	4.8	Cancerous prostate	76-44-1-E3-PU
ID267	est-not-ext	4.8	Hypertrophic prostate	77-3-4-H1-PU
ID268	est-not-ext	4.8	Cancerous prostate	76-45-2-C4-PU
ID269	est-not-ext	4.8	Prostate	60-12-1-D7-PU
ID270	est-not-ext	4.8	Normal prostate	78-46-2-B4-PU
ID271	est-not-ext	4.7	Prostate	60-12-3-A7-PU
ID272	est-not-ext	4.7	Normal prostate	78-24-3-A8-PU
ID273	est-not-ext	4.6	Hypertrophic prostate	77-17-3-A7-PU
ID274	est-not-ext	4.6	Hypertrophic prostate	77-10-1-F6-PU
ID275	est-not-ext	4.5	Prostate	60-13-1-E11-PU
ID276	est-not-ext	4.4	Normal prostate	78-24-3-C6-PU
ID277	est-not-ext	4.4	Cancerous prostate	76-23-1-B4-PU
ID278	est-not-ext	4.3	Hypertrophic prostate	77-9-1-E2-PU
ID279	est-not-ext	4.2	Normal prostate	78-4-4-B10-PU
ID280	est-not-ext	4.2	Normal prostate	78-30-2-C1-PU
ID281	est-not-ext	4.2	Normal prostate	78-38-2-E9-PU
ID282	est-not-ext	4.2	Normal prostate	78-8-2-F2-PU
ID283	est-not-ext	4.1	Cancerous prostate	76-20-3-H1-PU
ID284	est-not-ext	4.1	Cancerous prostate	76-14-1-B3-PU
ID285	est-not-ext	4.1	Normal prostate	78-18-4-D6-PU
ID286	est-not-ext	4	Hypertrophic prostate	77-11-4-B3-PU
ID287	est-not-ext	4	Normal prostate	78-16-2-C2-PU
ID288	est-not-ext	4	Hypertrophic prostate	77-38-2-G5-PU
ID289	est-not-ext	3.9	Normal prostate	78-25-1-H11-PU
ID290	est-not-ext	3.9	Hypertrophic prostate	77-12-3-H7-PU
ID291	est-not-ext	3.8	Cancerous prostate	76-21-4-A3-PU
ID292	est-not-ext	3.8	Normal prostate	78-41-1-C6-PU
ID293	est-not-ext	3.7	Cancerous prostate	76-5-2-H11-PU
ID294	est-not-ext	3.7	Cancerous prostate	76-8-4-D9-PU
ID295	est-not-ext	3.7	Cancerous prostate	76-18-2-D4-PU
ID296	est-not-ext	3.7	Prostate	60-12-3-G4-PU
ID297	est-not-ext	3.7	Hypertrophic prostate	77-20-2-E11-PU
ID298	est-not-ext	3.6	Cancerous prostate	76-1-2-G6-PU
ID299	est-not-ext	3.6	Normal prostate	78-8-3-F2-PU
ID300	est-not-ext	3.6	Normal prostate	78-12-4-E9-PU

SEQ. ID NO.	CATEGORY	VON HEIJNE SCORE	TISSUE SOURCE	INTERNAL DESIGNATION
ID301	est-not-ext	3.6	Hypertrophic prostate	77-15-2-E2-PU
ID302	est-not-ext	3.5	Cancerous prostate	76-7-3-A12-PU
ID303	est-not-ext	3.5	Normal prostate	78-22-3-E10-PU
ID304	est-not-ext	3.5	Hypertrophic prostate	77-2-3-E11-PU
ID305	est-not-ext	3.5	Normal prostate	78-29-1-B2-PU
ID306	ext-vrt-not-genomic	12	Normal prostate	78-47-2-C1-PU
ID307	ext-vrt-not-genomic	12	Normal prostate	78-43-4-G12-PU
ID308	ext-vrt-not-genomic	12	Hypertrophic prostate	77-38-1-A8-PU
ID309	ext-vrt-not-genomic	8.9	Normal prostate	78-45-4-F12-PU
ID310	ext-vrt-not-genomic	8.1	Normal prostate	78-35-3-D1-PU
ID311	ext-vrt-not-genomic	7.7	Normal prostate	78-10-1-H8-PU
ID312	ext-vrt-not-genomic	6.9	Cancerous prostate	76-43-1-E3-PU
ID313	ext-vrt-not-genomic	5.9	Normal prostate	78-29-2-C10-PU
ID314	ext-vrt-not-genomic	5.3	Hypertrophic prostate	77-38-3-B11-PU
ID315	ext-vrt-not-genomic	5.1	Normal prostate	78-36-4-A8-PU

TABLE III

SEQ. ID NO.	SIGNAL PEPTIDE
ID38	MVFVHLYLGNVLALLLFVHYSNG
ID39	MGMCFAAESDVQMFIAFLLCIFLICAALA
ID40	MAVRELCFSRQRQVLFLFLFWGVSLA
ID41	MRILQLILLALATGLVGG
ID42	MRILQLILLALATGLVGG
ID43	MRSCLWRCRHLSQGVQWSLLLAVLVFFLFA
ID44	MRILQXILLALATGLVGG
ID45	MLEECGAGVDLGFGGVKFASETPNLLWLLLKLVSTXWA
ID46	MIACSIRELHRCLLLALVAESSS
ID47	MGPPSLVLCLLSATVFS
ID48	MPGPRVWGKYLWRSPHSKGCPGAMWWLLLWGVLQX
ID49	MHRPEAMLLLLTLALLGGPTWX
ID50	MVSVSLALLSGWVGS
ID51	MHIFSICCMXSELHKMKSLSLQLASEKRSLVALVEEIVFLLLRVSPCLG
ID52	MKLWVSALLMAWFGVLS
ID53	MKVLISSLLLLLPLMLMSMVSS
ID54	MKVLISSLLLLLPLMLMSMVSS
ID55	MLLLLQLSLPSPTS
ID56	MLKMLSFKLLLLAVALG
ID57	MHRPEAMLLLLTLALLGXXXWA
ID58	MLKVSAVLCVCAAAWC
ID59	MKVGVLWLISFFTFTDG
ID60	MCIILLDLICLLFITA
ID61	MDCASISVKFTSMATMHDLSQFWASRGEVTNWWPVGQTSLPLFYLAFMVFGSFFPLISC
ID62	MTASPDYLVVLFGITAGATG
ID63	MVCVLVLAAAAGAVA
ID64	MKKTGDGGTLSTERIGGAALLSLLLKRMKMTLMIPLLLLTPITA
ID65	MELGCWTQLGLTFLQLLLISSLP
ID66	MRXKWKMGGMKYIFSLLFFLLLEGGXT
ID67	MRGATRVSIMLLLVTVSDC
ID68	MIAISAVSSALLFSLLCEAST
ID69	MIAISAVSSALLFSLLCEAST
ID70	MDPNGGCCTLLTLVLCVAVAYE
ID71	MEGEIYFQVFLSLFTFSTSLPSSLS
ID72	MYVVAMFGNCIVVFIVRTERSLHAPMYLFLCMLAAIDLALS
ID73	MRETXPLPKPLKDTAPSSHGVGSDSPSATRPWFLAPWCPGTQS
ID74	MDRPGSLSVFGSLPASLGTWLSSPAWLVDRPVRSAHPSANSTGVRMSVLVVLALRSLGRS
ID75	MHYFVAGKVILLFSYPSCCLC
ID76	MDLNSASTVVLQVLTQATS
ID77	MSSCNFTHATFVLIGIPGLEKAHFWVGFPLLSMYVVAMFGNCIVVFIVRTERSLHAPMYL FLCMLAAIDLALS
ID78	MYRLSLIAGPGSYPVLRWGVWDIPSSLVQVTYHQPNLTTNLDLPLFFSCSISATHS
ID79	MLVDGPSERPALCFLLLAVAMSFF
ID80	MPCSLTWRLPPRTCQXXGLXKSXLXXLLTPPPSYG
ID81	MVXWLVLFALQIYSYXSTRDQPASRXRLLFLFLTSIAEXCS
ID82	MARHGLPLLXXXSLPVGA
ID83	MVHLRTGLMLMSADRLRTLYYTVTILYILWYCSVCSS
ID84	MGILSTVTALTFARA
ID85	MELGCWTQLGLTFLQXLLISSLX
ID86	MELLRVCSFFLLCXSVFTDCKG
ID87	MIVRPRLNLTWFLLLPPGQCRA

SEQ. ID NO.	SIGNAL PEPTIDE
ID88	MQFLFKMVALCCCLWKISG
ID89	MLKVSAVLCVCAAAXXSQSLX
ID90	MSMQFLFKMVALCCCLWKISG
ID91	MAQHLWILLGSLSCRTS
ID92	MNKEXVSXERXAQVRLYLFSGFWTFXLG
ID93	MVLWRAKIXRNVPVTLSEENRSEGKVGFQAYKNYFRAGAHWIVFIFLILLNTAA
ID94	MLLXFFTSVLWLTSPSQP
ID95	MELISPTVIIILGCLALFLLLQ
ID96	MHGFEIISLKEESPLGKVSQGPLFNVTSGSSSPVTWLGLLSFQNLHC
ID97	MTWVRHAPGKSLEWVATVTDGGDKTFYAASVKGRFNVSRDNSKNTLFLHLSGLSAA
ID98	MLTSFFSLTANCQS
ID99	MLLCLLTPLFFMXPTGFS
ID100	MDDDYEAYHSLFLSLLGLCPS
ID101	MEWGKQWLVWLLLGHMVVS
ID102	MRRGKRLLESQSSSPKACLQLGFETELTQGVLWILVIQA
ID103	MVAATEAALLESVVWLPCHG
ID104	MSWNPSVSLPLLSSWGSTA
ID105	MKRIQGILFLILLSLHLERRWT
ID106	MVQRLWVSRLLRHRKAQLXLXNLLTFGLEVCLAAG
ID107	MAAGVPFALVTSCSSVFS
ID108	MTVFLXFCFPRCHS
ID109	MXPNNFWQKLGRKKPRIFTCTQSSTGEAAVKAENLILLEVFVWNGLQG
ID110	MFRSDRMWXCHWKWKPSPLLFLFALYIMCVPHSVWG
ID111	MTQRSIAGPICNLKFVTLLVALSSELPFLGA
ID112	MIIPLLLLLRSACN
ID113	MXSPLPVLLLSXNLNLIIQ
ID114	MLMCKMLKSQKNCQENXXIKIILFLKPMCSPQYLLTFLVFTXKLSS
ID115	MKKKSSPNQYLHSSLHXIRLFSFLHFSEEGVLLLAIDLKIIVILHCAASIIS
ID116	MFSCFFSTSLATSVSLEAQSCFA
ID117	MHHGLTPLLLGVHEQKQQVVKFLIKKKANLNALDRYGRTALILAVCCGSA
ID118	MSPCIYFFACFQALTSS
ID119	MAEEMESSLEAXFSSSGAVSGASGFLPPARS
ID120	MAEEMESSLEASFSSSGAVSGASGFLPPARS
ID121	MLVLGSPLLGPLLWHLSLILLKPLCLP
ID122	MHLLDLESMGKSSDGKSYVITGSWNPKSPHFQVVNEETPKDKVLFMTTAVDLVIT
ID123	MENLKDFYVLFVFSSIPLTFL
ID124	MPQYCLSIFSLVLPVCRM
ID125	MVAPVLETSHVFCCPNRVRGVLNWSSGPRGLLAFGTSCSVVLY
ID126	MPIIDQVNPELHDFMQSAEVGTIFALSWLITWFGHXLS
ID127	METXCPCCCCPCXGXGSLXXKPVYELQVQKSVTVQEGLCVLVPCSXSXX
ID128	MSPCIYFFACFXXLTSS
ID129	MGRGERRHYWGPKLVLKCLSFSXPSLP
ID130	MSQDGGXGELKHMVMSFRVSELQVLLGFAGRNKSGRKHELLAKALHLLKSSC
ID131	MHHRMNEMNLSPVGMEQLTSSSVSNALPVSGSHLGLAASPTHSAIPAPGLPVAIPNLGPS LSSLPSALS
ID132	MLHSDNIWNLFSLFSTSTT
ID133	MQPASPPARWSFHSAAGWSGGGQA
ID134	MCFSFLLAGSISHMFSQA
ID135	MYGFIIGLSILFHCSVCLFLC
ID136	MSFGXILTFRVSLLGCXLAININT
ID137	MAVYVGMLRLGRLCAGSSGVXG
ID138	MFNTIYLVISLVSIFFFWEVTNA
ID139	MALPPKGCGSLPLTTGSSWSLS

SEQ. ID NO.	SIGNAL PEPTIDE
ID140	MFVFLSWASFLAPLLR
ID141	MXMKSANKITLLXHHLLSCSPLXPLGKS
ID142	MCNYNIYVLYNIGYLYHPKSFLLLFIVIPQTP
ID143	MAVAMVKLCERAGLPLLAAPLLRSLLP
ID144	MLNVVRALRXPQWCAEYCLSIHYQHGGVICTQVHKQTVVQLALRVADEMDVNIGHEVGYV IPFENCCTNETILRYCTDDMLQREMMSNPFLGSYGVIILDDIHERSIATDVLLGLLKDVLLA
ID145	MHAGLERXSXQKALAGLCIGSTSYVHG
ID146	MLNGPFQHRNSRIMTHRSAEKTLLGSLSLWRWSAM
ID147	MRVKDPTKALPEKAKRSKRPTVPHDEDSSDDIAVGLTCQHVSHA
ID148	MPQKGLGLLGILSGDFSLLALSMLKGTG
ID149	MAMWNRPXXXLPQQPLXAEPTAEGEPHLPTGRXXTEANRFAYAALCGISLSQLFP
ID150	MLCFGDLLLSPWVTVPVWS
ID151	MQENAHNLRLFKCLLIYFLGLAADTYF
ID152	MHTCSLPCLLFAQLLEFCSFPPDVPHNCAPIVSVRPPNIVAAFEGCSVATALFPPLCIS
ID153	MQQRGAAGSRGCALFPLLGVLFFQGVYI
ID154	MXXSIFISEKYGLCPSKTPIMKMLPSLILNRSLPTASSS
ID155	MAFDVSCFFWVVLFSAGCKV
ID156	MEVAANCSLRVKRPLLDPRFEGYKXSLEPLPCYQLELDAAVAXVKLRDDQYTLEHMHAFG MYNYLHCDSWYQDSVYYIDTLGRIMNLTVMLDTAXG
ID157	MNVGTAHXXVNPNTRVMNSRGIWLSYVLAIGLLHIVLLS
ID158	MENFNMYKNKSWWTLLSSSPSFM
ID159	MNVGTXHSEVNPNTRVMNSRGIWLSYVLAIGLLHIVLLS
ID160	MAAASAVSVLLVAA
ID161	MAYSKASGSPVLSQAVPGENASHRRGSADLGSGSGLSWARLSQS
ID162	MKPRRNLEEDDYLHKDTGETSMLKRPVLLHLHQTAHA
ID163	MIICYDIPCAHMLVCPTIG
ID164	MYSSEDSTLASVPPAATFG
ID165	MGEDPXQPRKYKKXKXELQGDXPPSSPTNDPTVKYETQPRFITATGGTLHMYQLEGLNWL RFSWA
ID166	MFYVAMTKTHKRIRSLCNIHHGLFQFTQQLLGCLQCCWLQSG
ID167	MVSPKDLPLVLLQDIKVPSSMTGSHAGNPHIERNDLPRHGSPQFFTGXTCASXNPSQCLA
ID168	MEFXSLFCLYFSCFL
ID169	MALHFQSLAELEXLCTHLYIGTDLTQRIEAEKALLELIDSPECLS
ID170	MRTLFGAVRAPFSSLTLLLITPSPSPL
ID171	MRHSLLKGISAQIVSAADKVDAGLPTAIAVSSLIAVGTSHG
ID172	MTLSCFIFFYISSLC
ID173	MILCFLLPHHRLQEA
ID174	MFSLFALNMPLGFC
ID175	MASSPGVAMHSLWATIHTSVWGVLPPPACSA
ID176	MSQEGAVPASAVPLEELSSWPEELCRRELPSVLPRLLSLSQHSES
ID177	MTRECPSPAPGPGAPLSGSVLAEAAVVFAVVLSIHA
ID178	MQELHLLWWALLLGLAQA
ID179	MGRQALLLLALCATGAQG
ID180	MGPSTPLLILFLLSWSGPLQG
ID181	MSCRELTHRPCSPHLLLLCPLSRGCCP
ID182	MGWTMRLVTAALLLGLMMVVTG
ID183	MKFLIFAFFGGVHLLSLCSGKVYA
ID184	MQCFSFIKTMMILFNLLIFLCGAALLAVG
ID185	MWAFSELPMPLLINLIVSLLGFVATVTL
ID186	MASSNTVLMRLVASAYSIA
ID187	MKFLIFAFFGGVHLLSLCSGKAIC
ID188	MADTTPNGPQGAGAVQFMMTNKLDTAMWLSRLFTVYCSALXVLPLLGLHEA
ID189	MRFRHFXKXIGXVLVLSVVXXAMA

SEQ. ID NO.	SIGNAL PEPTIDE
ID190	MELGSCLEGGREAAEEEGEPEVKKRRLLCVEFASVASCDA
ID191	MASPFSGALQLTDLDDFIGPSQECIKPVKVEKRAGSGVAKIRIEDDGSYFQINQDGXTRRLE KAKVSLNYCXACSGCITSAETVLITQQSHEELKKVLDANKMAAPSQQRLVVVSVSPQSRA
ID192	MGPVPTAVAGAGSRLVKPSQTLSLTCAVSGGSLVAELLLGAGSG
ID193	MESGGRPSLCQFILLGTTSVVTA
ID194	MQVCRCIYIICFXLPPLFS
ID195	MAQRLLLRFLASVIS
ID196	MLFIFNFLFSPLPTPALICILTFGAAIFLWLITRPQPVLP
ID197	MYPKWEAPVTFCQLKREKDPPHPAHSPFLQPRFSHMLQLLPSKALC
ID198	MALYQRWRCLRLQGLQACRLHTAVVSTPPRWLAERLGLFEELWA
ID199	MGVPRPQPWAXGLLLFLLPGSLG
ID200	MAAAVPKRMRGPAQAKLLPGSAIQALVGLARPLVLALXLVSAALS
ID201	MWLWEDQGGLLGPFSFLLLVLLLVTRXRS
ID202	MNWELLLWLLVLCALLLLLVQLLRFLRA
ID203	MEKIPVSAFLLLVALSYTLA
ID204	MSNYTDAESSFSKQEIIRVAMEKIPVSAFLLLVALSYTLA
ID205	MQFXTWATSSSQPALWSLLLVSWAAMVLRLRSKCALVTFFFILLLIFIAEVAA
ID206	MNWELLLWLLVLCALLLLLVHLLRFLRA
ID207	MTTFLPVPQMMAGFSFGTFGNPPMESPSAWQTIHQPFIVSCLTLWSPGCWP
ID208	MASKGMRHFCLISEQLVXFSLLATAILG
ID209	MAAAAWLQVLPVILLLLGAHP
ID210	MASPRTVTIVALSVALGLFFVFMGTIKLTPRLSKDAYSEMKRAXKSYVRALPLLKKMGIN SILLRKSIGALEVACGIVMTLVPGRPKDVANFFLLLLVLAVLFFHQLVG
ID211	MPNLSFGGLDTNQMRVNFLSVDVCKLLLLCALHSHIYC
ID212	MGPPMLQEISNLFLILLMMGAIFTLAALKESLSTCIPAIVCLXXLLLLNVGQLLA
ID213	MXXFTDPSSVNEKKRREREERQNIVLWRQPLITLQYFSLEILVILKEWTSKLWHRXXIVV XFLLLLAXLIA
ID214	MPLLRGLLWXQVLCA
ID215	MKLLSLVAVVGCLLVPPAEA
ID216	MPALLPVASRLLLLPRVLLTMASG
ID217	MCLLLGATGVGKTLLVKRLQEVSSRDGKGDLGEPPPTRPTVGTNLTDIVAQRKITIRELG GCMGPIWSSYYGNCRSLLFVMDASDPTQLSAXXVQLLGLLSAEQLAEA
ID218	MELPAVNLESDSPRSLAADNLGLHCILRLLCLGQLHHPGLG
ID219	MAFLRKVYSILSLQVLLTTVTSTVFLYFESVRTFVHESPALILLFALGSLG
ID220	MYTYGNKQHNSPTWDDPTLAIALAANAWA
ID221	MQQIFIQQCRELNFWSREPWILVLALPLTVWP
ID222	MKAVLLALLMAGLAL
ID223	MGLQACLLGLFALILS
ID224	MRPGQVSLLGPDAVSVLGSGLGLSPGTSS
ID225	MINPSVPSKSNSHPFLSTVMFTSASLLLPMSTG
ID226	MSEKEXNFPPLPKFIPVKPCFYQNFSDEIPVEHQVLVKRIYRLWMFYCATLGVNLIACLA WWIGGGSG
ID227	MNPTKLILKTILRLYFFLQLAHS
ID228	MASSSPDSPCSXXCFVSVPPASA
ID229	MXPVLAALAHVLCPYMAPGLCREPIRXLIAXLEPPGAMA
ID230	MNNLNDPPNWNIRPNSRADGGDGSRWNYALLVPMLGLAAFRWIWS
ID231	MLLLFLAALCSLFFFLSLQ
ID232	MLFLGKVLIVCSTGLAGIMLLNYQQDYTVWVLPLIIVCLFAFLVAHC
ID233	MQGIPILTPVTTQSIAISIVLTVQGLLLLVHSFWFTVC
ID234	MQNFCHHLAICTVILFCVLLSLRPHTS
ID235	MPSFSKDLLTVPKLGTGHXXGXGSYDXALXLLLKCLWSNVVPECTMASSNTVLMRLVASA YSIA
ID236	MRGAHLTALEMLXAFASHIXA

SEQ. ID NO.	SIGNAL PEPTIDE
ID237	MEVGLPAITLFLTSASSPVVATTMDQEPVGGVERGEAVAASGXAAAAAFGESAGQMSNERGFENVELGVIGKKKKVPRRVIHFVSGETMEEYSTDEDXVDGLEKXMFCLLLIRQNLPGVPTYGFTCFGLLHQLSQCVTS
ID238	MKELERQQKEVEERPEKDFTEKGSRNMPGLSAATLASLGGTSS
ID239	MSMGFMMLVLVILCIVTVCVT
ID240	MMELXLKXXTKXEXESACTEAYSQSDEQYACHLGCQNQLPFAELRQEQLMSLMPKMHLLFPLTLVRSFWS
ID241	MVSNASETSCLGLILLFASHLINQ
ID242	MPRKRKCDLRAVRVGLLLGGGGVYGSRFRFTFPGCRALSPWRVRXQRRRCEMSTMFADTLLIVFISVCTALLA
ID243	MGMWSIGAGALGAAALALLLANT
ID244	MDVAFLEXLIKDDIERGRLPLLLVANAGTAA
ID245	MRTLFNLLWLALACSP
ID246	MNAQPGLXLDCITRFLTXGQFICLQWALPHSEA
ID247	MGKEWGWQEMENGGAAPAWGAGPPVHPAPPPVEKTLSWGCGFGLHSGFGGSGGGVGLCRLLCLVRLFCC
ID248	MAAPSGGWNGVGASLWAALLLTATVRLSA
ID249	MIAIYGKNFCVSAKNAFMLLMRNIVRVVVLDKVTDLLLFFGKLLVVGG
ID250	MERNCKGSFGVIKEGDTDTXETKARRTVWEPRGRYSFRXTPRPAYPVEQCGFARRALELLEIRKHSPEVCEPPNIPVTSVLELIVASVCQS
ID251	MFVEYRKQLKLLLDRLAQVSPELLLASVRRVFSSTLQNWQTTRFMEVEVAIRLLYMLAEALPVSHG
ID252	MLLGTSNIIIFLIQWHGSVFQ
ID253	MXNRFATAFVXACVLSLIST
ID254	MSLTSGFLRVSQG
ID255	MANFKGHALPGSFFLIIGLCWSVKYPLKYFSHTRKNSPLHYYQRLEIVEAAIRTLFSVTGILA
ID256	MQDTGSVVPLHWFGFGYAALVASGGIIGYVKAGSVPSLAAGLLFGSLAGLGA
ID257	MEXGLKSADPRDGTGYTXXXXYCCALLTSLXCIWG
ID258	MASPSRRLQTKPVITCFKSVLLIYTFIFWITGVILLAVGIWG
ID259	MFSRELAPTRIGGASSGSRSGGTLISTAPLTTRVLNPTAQCFCLDCTLRRMQTHLSVSLLPCAGAWS
ID260	MSMAVETFGFFMATVGLLMLGVTLPNSYW
ID261	MEKIPVSXFLXLXXLSXXWP
ID262	MHSAEEPLXLAALRGARGHLPCGSRHHVGSLAPASVPAPGACLWVCEWETLLPGLILERPLVPSAEA
ID263	MAGQFRSYVWDPLLILSQIVLMQTVYYGSLGLWLALVDGLVRX
ID264	MAPKVFRQYWDIPDGTDCHRKAYSTTSIASVAGLTAAAYRVTLNPPGTFLEGVAKVGQYTFTAAAVGAVFGLTTCISA
ID265	MAAAAWLQVLPVILLLLG
ID266	MEIYFIFCIIVPIAAATVYKSWCLLLILDMNVLYTDA
ID267	MSRYTSPVNPAVFPHLTVVLLAIGMFFTAWF
ID268	MRLAAEAHPGRTHTLFRRLKPFLMLSSSLPLLIWL
ID269	MLEHLXSLPTQMDYKGQKLAXQMFQGIILFSAIVGFIYG
ID270	MEYSKVLFCSFSNVLG
ID271	MASKIGSRRWMLQLIMQLGSVLLTRC
ID272	MEHYRKAGSVELPAPSPMPQLPPDTLEMRVRDGSKIRNLLGLALGRLEGGSA
ID273	MNALMVLFNVTVVLIALTCLDGTTVS
ID274	MNWSIFEGLLSGVNKYSTAFGRIWLSLVFIFRVLVYLVTAERVWS
ID275	MIISLFIYIFXTCSNT
ID276	MFRLNSLSALAELAVG
ID277	MTAGTLRTWLCCAGSWA
ID278	MLGRPCFHSPQRLLVILCVSVKAG
ID279	MDEARDNACNDMGKMLQFVLPVATQIQQ

SEQ. ID NO.	SIGNAL PEPTIDE
ID280	MSPISIRELCALGSAPSSMWA
ID281	MTDLLSASPWALT
ID282	MSWSGLLHGLNTSLTCGPALVPRLWA
ID283	MADVINVSVNLEAFSQAISAIQA
ID284	MNVIDHVRDMAAAGLHSNVRLLSSLLLTMSNN
ID285	MTSACLAWTAVRPSAC
ID286	MNGSRTLTHSISDGQLQGGQSNSELFQQEXQTAPAQVPQGFNVFGMSSSSGASNS
ID287	MLGFFLFLSFVLMYDG
ID288	MMEERANLMHMMKLSIKVLLQSALSLG
ID289	MELEXIVSAALLAFVQT
ID290	MLRQIIGQAKKHPSLIPLFXFIGTGA
ID291	MVKETQYYDILGVKPSASPERSRRPIGSWRSSTTRTRTRTRMRARSLNSYPRHMKCFQIQRK GMFMTKAESRQXKKEAQAAPASLHPWTSLTCSLVVVDG
ID292	MANLFIRKMVNPLLYLSRHTVKPRALSTXLFGSIRG
ID293	MAAAAASRGXGAKLGLRXIRIHLCQRSPGSQG
ID294	MFPSCYLCYSLCGSILLSIFSAYNRLSLMLRIALTLIPSMLSRA
ID295	MSTQXGLSMHAHPQAYTPFIYLHARKRRGEIGDADSRFNDRYAHKSAQLXFLYFVCCIFQ
ID296	MKHFQDLPSSCSCSLISFTRG
ID297	MSQRSLCMDTSLDVYRXLIELNYLGTVSLTKCVLPHMIERKXXKIVTVNSILGIISVPLSIG
ID298	MGGSGSRLSKELLAEYQDLTFLTKQEILLAHRRFCELLPQEQRXXSRHFGHKCPSSRFSA FQSSRPTPSRSESAGSSPHPQPKTALALRTSWISSVCS
ID299	MWRLLARASAPLLRVPLSDSWALLPASA
ID300	MADHVQSLAQLENLCKQLYETTDTXXRSSXAEKALVEFTNSPDCLSKCQLLLERGSSSYS QLLAATCLTKLVSRTNNPLPLEQRIDIRNYVLNXLATRPKLATFVTQALIQXYA
ID301	MAYHGLTVPLIVMSVFWGFVGFLVPWFIPKGPNRGVIITMLVTCSVCCYLFWLIA
ID302	MSTGQLYRMEDIGRFHSQQPGSLTPSSPTVGEIIYNNTRNTLGWIGGILMGSFQGTIA
ID303	MGWQRWWCFHLQAEASA
ID304	MSVIFFACVVRVRDG
ID305	MAVTALAAXTWLGVWG
ID306	MSLSAFTLFLALIGGTSG
ID307	MSLSAFTLFLALIGGTSG
ID308	MSLSAFTLFLALIGGTSG
ID309	MVELMFPLLLLLLPFLLYMA
ID310	MWLLYLLVPALFCRA
ID311	MKQILHPALETTAMTLFPVLLFLVAGLLPSFP
ID312	MLKALFLTMLTLALVKS
ID313	MEKNPLAAPLLILWFHLDCVSS
ID314	MRVVTIVILLCFCKA
ID315	MDQFPESVTENFEYDDLAEACYIGDIVVFGTVFLSIFYSVIFAIGLVGNLLVVFALTNSK KPKSVTDIYLLNLALSDLLFVATLPFWTHY

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

TABLE IV

Minimum signal peptide score	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
3.5	2674	947	599	23	150
4	2278	784	499	23	126
4.5	1943	647	425	22	112
5	1657	523	353	21	96
5.5	1417	419	307	19	80
6	1190	340	238	18	68
6.5	1035	280	186	18	60
7	893	219	161	15	48
7.5	753	173	132	12	36
8	636	133	101	11	29
8.5	543	104	83	8	26
9	456	81	63	6	24
9.5	364	57	48	6	18
10	303	47	35	6	15

TABLE V

Tissue	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
Brain	329	131	75	3	24
Cancerous prostate	134	40	37	1	6
Cerebellum	17	9	1	0	6
Colon	21	11	4	0	0
Dystrophic muscle	41	18	8	0	1
Fetal brain	70	37	16	0	1
Fetal kidney	227	116	46	1	19
Fetal liver	13	7	2	0	0
Heart	30	15	7	0	1
Hypertrophic prostate	86	23	22	2	2
Kidney	10	7	3	0	0
Large intestine	21	8	4	0	1
Liver	23	9	6	0	0
Lung	24	12	4	0	1
Lung (cells)	57	38	6	0	4
Lymph ganglia	163	60	23	2	12
Lymphocytes	23	6	4	0	2
Muscle	33	16	6	0	4
Normal prostate	181	61	45	7	11
Ovary	90	57	12	1	2
Pancreas	48	11	6	0	1
Placenta	24	5	1	0	0
Prostate	34	16	4	0	2
Spleen	56	28	10	0	1
Substantia nigra	108	47	27	1	6
Surrenals	15	3	3	1	0
Testis	131	68	25	1	8
Thyroid	17	8	2	0	2
Umbilical cord	55	17	12	1	3
Uterus	28	15	3	0	2
Non tissue-specific	568	48	177	2	28
Total	2677	947	601	23	150

TABLE VI

Description of Transcription Factor Binding Sites present on promoters isolated from SignalTag sequences

Promoter sequence P13H2 (646 bp):

Matrix	Position	Orientation	Score	Length	Sequence
CMYB_01	-502	+	0.983	9	TGTCAGTTG
MYOD_Q6	-501	-	0.961	10	CCCAACTGAC
S8_01	-444	-	0.960	11	AATAGAATTAG
S8_01	-425	+	0.966	11	AACTAAATTAG
DELTAEF1_01	-390	-	0.960	11	GCACACCTCAG
GATA_C	-364	-	0.964	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHAE47_01	-235	+	0.973	16	CATAACAGATGGTAAG
TAL1BETAE47_01	-235	+	0.983	16	CATAACAGATGGTAAG
TAL1BETAITF2_01	-235	+	0.978	16	CATAACAGATGGTAAG
MYOD_Q6	-232	-	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	TCAAGATAAAGTA
IK1_01	-126	+	0.963	13	AGTTGGGAATTCC
IK2_01	-126	+	0.985	12	AGTTGGGAATTC
CREL_01	-123	+	0.962	10	TGGGAATTCC
GATA1_02	-96	+	0.950	14	TCAGTGATATGGCA
SRY_02	-41	-	0.951	12	TAAAACAAAACA
E2F_02	-33	+	0.957	8	TTTAGCGC
MZF1_01	-5	-	0.975	8	TGAGGGGA

Promoter sequence P15B4 (861bp) :

Matrix	Position	Orientation	Score	Length	Sequence
NFY_Q6	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-682	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCCTGGAA
STAT_01	-673	-	0.951	9	TTCCAGGAA
MZF1_01	-556	-	0.956	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	-	0.955	12	GAAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

Promoter sequence P29B6 (555 bp) :

Matrix	Position	Orientation	Score	Length	Sequence
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	-	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	-	0.972	12	CAGCACGTGAGT
USF_C	-307	+	0.997	8	TCACGTGC
USF_C	-307	-	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

TABLE VII

CLAIMS

1. A purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 38-315 or comprising a sequence complementary thereto.

2. The nucleic acid of Claim 1, wherein said nucleic acid is recombinant.

3. A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary thereto.

4. A purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-315 or one of the sequences complementary thereto.

5. The nucleic acid of Claim 4, wherein said nucleic acid is recombinant.

6. A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-315 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315.

7. The nucleic acid of Claim 6, wherein said nucleic acid is recombinant.

8. A purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-315.

9. A purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-315 or having a sequence complementary thereto.

10. A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 38-315 which encode a signal peptide.

11. A purified or isolated polypeptides comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-315.

12. A vector encoding a fusion protein comprising a polypeptide and a signal peptide, said vector comprising a first nucleic acid encoding a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-315 operably linked to a second nucleic acid encoding a polypeptide.

13. A method of directing the extracellular secretion of a polypeptide or the insertion of a polypetide into the membrane comprising the steps of:

obtaining a vector according to Claim 12; and

introducing said vector into a host cell such that said fusion protein is secreted into the extracellular environment of said host cell or inserted into the membrane of said host cell.

14. A method of importing a polypeptide into a cell comprising contacting said cell with a fusion protein comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-315 operably linked to said polypeptide.

15. A method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-315, comprising the steps of:

obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-315;

contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-315 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA;

identifying a cDNA which hybridizes to said detectable probe; and

isolating said cDNA which hybridizes to said probe.

16. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 15.

17. The cDNA of Claim 16 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

18. A method of making a cDNA comprising one of the sequences of SEQ ID NOs: 38-315, comprising the steps of:

contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;

hybridizing said first primer to said polyA tail;

reverse transcribing said mRNA to make a first cDNA strand;

making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-315; and

isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

19. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 18.

20. The cDNA of Claim 19 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

21. The method of Claim 18, wherein the second cDNA strand is made by:

contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-315 and a third primer having a sequence therein which is included within the sequence of said first primer;

performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product;

contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NO:s 38-315, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and

performing a second polymerase chain reaction, thereby generating a second PCR product.

22. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-315, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 21.

23. The cDNA of Claim 22 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-315.

24. The method of Claim 18 wherein the second cDNA strand is made by:

contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-315;

hybridizing said second primer to said first strand cDNA; and

extending said hybridized second primer to generate said second cDNA strand.

25. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-315 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 24.

26. The cDNA of Claim 25, wherein said cDNA comprises the full protein coding sequence partially included in of one of the sequences of SEQ ID NOs: 38-315.

27. A method of making a protein comprising one of the sequences of SEQ ID NO: 316-593, comprising the steps of:

obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NO: 38-315;

inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;

introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and

isolating said protein.

28. An isolated protein obtainable by the method of Claim 27.

29. A method of obtaining a promoter DNA comprising the steps of:

obtaining DNAs located upstream of the nucleic acids of SEQ ID NO: 38-315 or the sequences complementary thereto;

screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and

isolating said DNA comprising said identified promoter.

30. The method of Claim 29, wherein said obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NO: 38-315 or sequences complementary thereto.

31. The method of Claim 30, wherein said screening step comprises inserting said upstream sequences into a promoter reporter vector.

32. The method of Claim 30, wherein said screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

33. An isolated promoter obtainable by the method of Claim 32.

34. An isolated or purified protein comprising one of the sequences of SEQ ID NO: 316-593.

35. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the improvement comprising inclusion in said array of at least one of the sequences of SEQ ID NOs: 38-315, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-315, or a fragment thereof of at least 15 consecutive nucleotides.

36. The array of Claim 35 including therein at least two of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides.

37. The array of Claim 35 including therein at least five of the sequences of SEQ ID NOs: 38-315, the sequences complementary to the sequences of SEQ ID NOs: 38-315, or fragments thereof of at least 15 consecutive nucleotides.

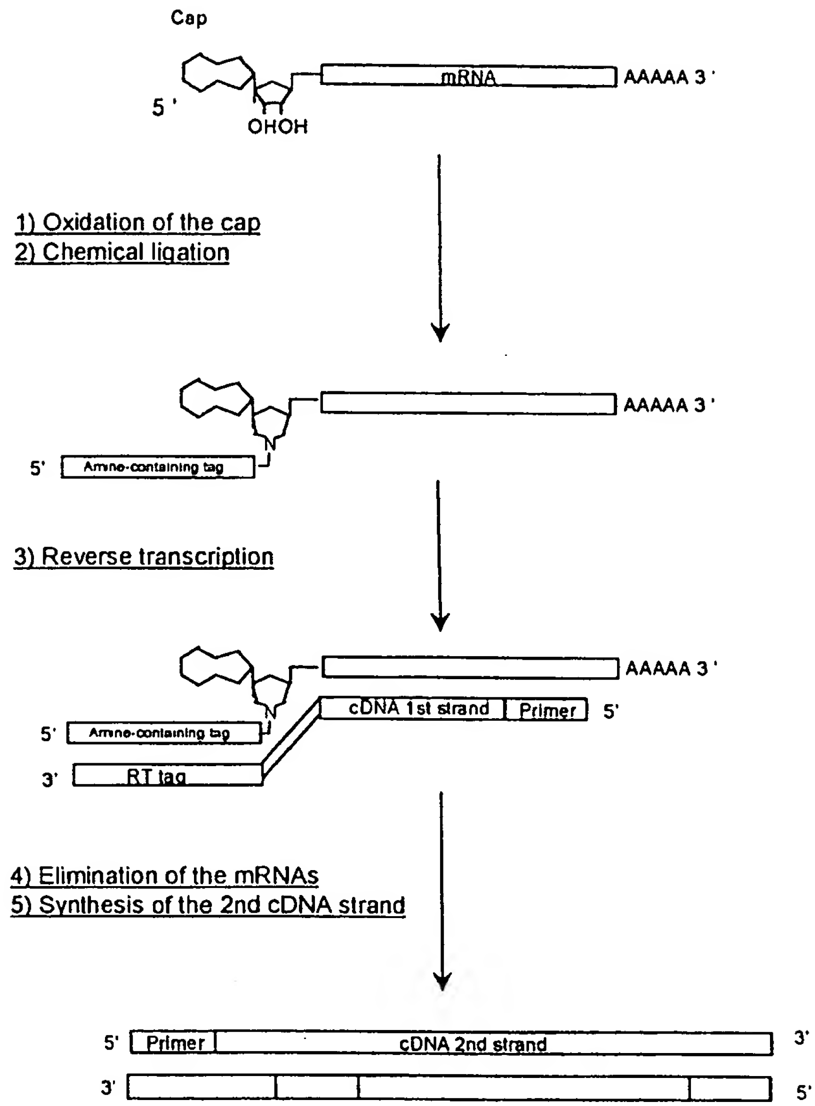

Cap
5'
OHOH
mRNA
AAAAA 3'
1) Oxidation of the cap
2) Chemical ligation
AAAAA 3'
5'
Amine-containing tag
3) Reverse transcription
AAAAA 3'
cDNA 1st strand
Primer
5'
5'
Amine-containing tag
3'
RT tag
4) Elimination of the mRNAs
5) Synthesis of the 2nd cDNA strand
5'
Primer
cDNA 2nd strand
3'
3'
5'

Figure 1

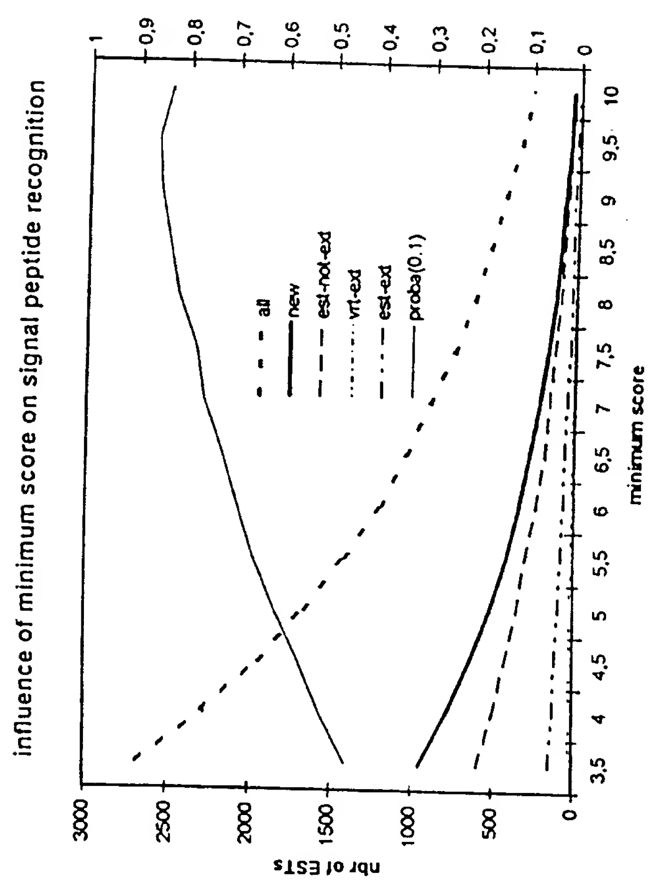

influence of minimum score on signal peptide recognition
nbr of ESTs
minimum score
all
new
est-not-ext
vrt-ext
est-ext
proba(0.1)
0
500
1000
1500
2000
2500
3000
3,5
4
4,5
5
5,5
6
6,5
7
7,5
8
8,5
9
9,5
10
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1

Figure 2

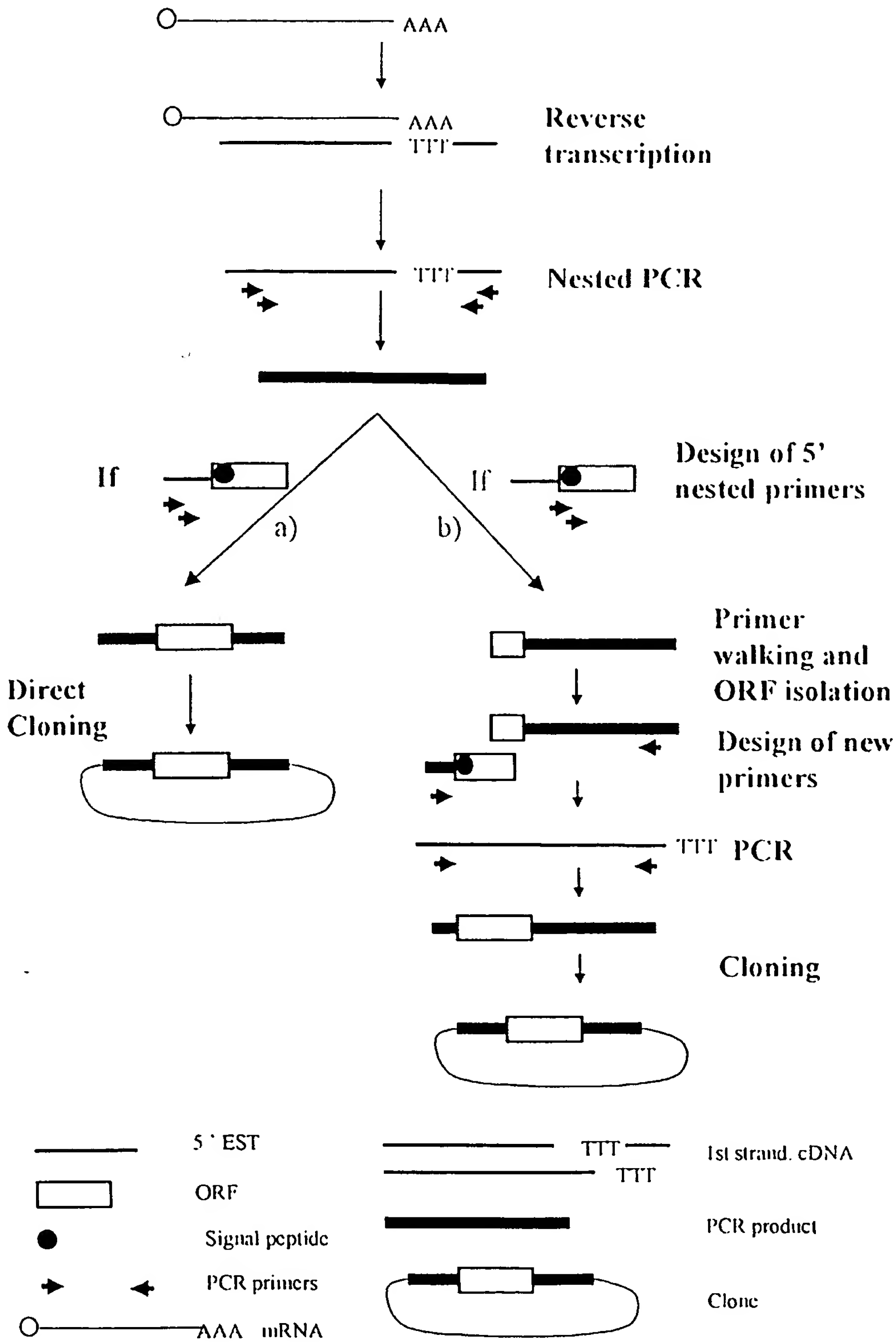

AAA
AAA
TTT
Reverse transcription
TTT
Nested PCR
If
If
a)
b)
Design of 5' nested primers
Direct Cloning
Primer walking and ORF isolation
Design of new primers
TTT
PCR
Cloning
5 ' EST
ORF
Signal peptide
PCR primers
AAA mRNA
TTT
TTT
1st strand. cDNA
PCR product
Clone

Figure 3

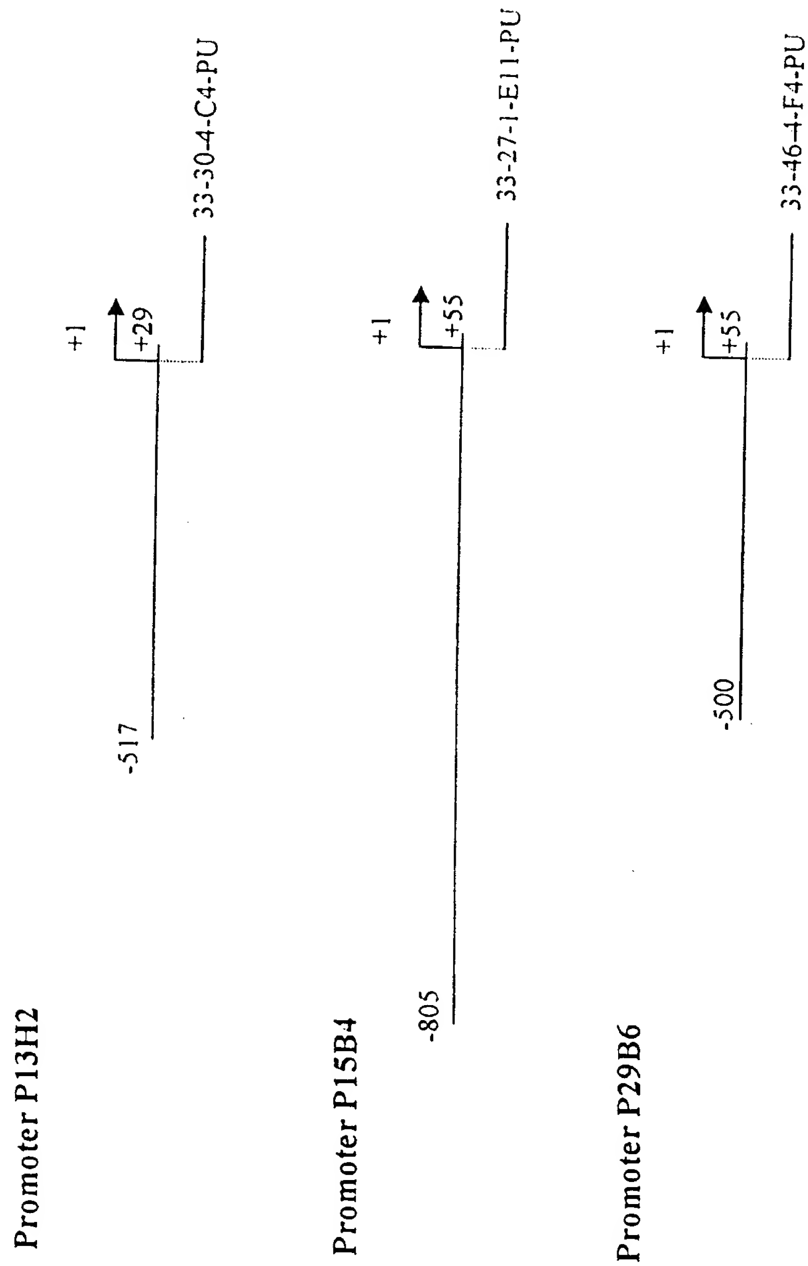

Promoter P13H2
-517
+1
+29
33-30-4-C4-PU
Promoter P15B4
-805
+1
+55
33-27-1-E11-PU
Promoter P29B6
-500
+1
+55
33-46-4-F4-PU

Figure 4

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME : GENSET SA
(B) STREET :24, RUE ROYALE
(C) CITY: PARIS
(E) COUNTRY : FRANCE
(F) POSTAL CODE (ZIP) : 75008

(ii) TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS EXPRESSED IN PROSTATE

(iii) NUMBER OF SEQUENCES: 593

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy Disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: Win95
(D) SOFTWARE: Word

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(ix) FEATURE:
(A) NAME/KEY: Cap
(B) LOCATION: 1
(D) OTHER INFORMATION: m7Gppp added to 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCAUCCUAC UCCCAUCCAA UUCCACCCUA ACUCCUCCCA UCUCCAC 47

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCAUCCUACU CCCAUCCAAU UCCACCCUAA CUCCUCCCAU CUCCAC 46

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCAAGAATT CGCACGAGAC CATTA 25

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TAATGGTCTC GTGCGAATTC TTGAT 25

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCGACAAGAC CAACGTCAAG GCCGC 25

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TCACCAGCAG GCAGTGGCTT AGGAG 25

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AGTGATTCCT GCTACTTTGG ATGGC 25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCTTGGTCTT GTTCTGGAGT TTAGA 25

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TCCAGAATGG GAGACAAGCC AATTT 25

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AGGGAGGAGG AAACAGCGTG AGTCC 25

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATGGGAAAGG AAAAGACTCA TATCA 25

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AGCAGCAACA ATCAGGACAG CACAG 25

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATCAAGAATT CGCACGAGAC CATTA 25

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATCGTTGAGA CTCGTACCAG CAGAGTCACG AGAGAGACTA CACGGTACTG GTTTTTTTTT 60

TTTTTVN 67

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCAGCAGAGT CACGAGAGAG ACTACACGG 29

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CACGAGAGAG ACTACACGGT ACTGG 25

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 526 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: complement(261..376)
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 166..281
id N70479
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: complement(380..486)
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 54..160
id N70479
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: complement(110..145)
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 94
region 403..438
id N70479
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: complement(196..229)
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 94
region 315..348
id N70479
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 90..140
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.2
seq LLLITAILAVAVG/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AATATRARAC AGCTACAATA TTCCAGGGCC ARTCACTTGC CATTTCTCAT AACAGCGTCA 60

GAGAGAAAGA ACTGACTGAR ACGTTTGAG ATG AAG AAA GTT CTC CTC CTG ATC 113

```
                                Met Lys Lys Val Leu Leu Leu Ile
                                        -15                 -10

ACA GCC ATC TTG GCA GTG GCT GTW GGT TTC CCA GTC TCT CAA GAC CAG      161
Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln
                -5                  1                   5

GAA CGA GAA AAA AGA AGT ATC AGT GAC AGC GAT GAA TTA GCT TCA GGR      209
Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly
        10                  15                  20

WTT TTT GTG TTC CCT TAC CCA TAT CCA TTT CGC CCA CTT CCA CCA ATT      257
Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile
    25                  30                  35

CCA TTT CCA AGA TTT CCA TGG TTT AGA CGT AAN TTT CCT ATT CCA ATA      305
Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile
40                  45                  50                  55

CCT GAA TCT GCC CCT ACA ACT CCC CTT CCT AGC GAA AAG TAAACAARAA       354
Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys
                60                  65

GGAAAAGTCA CRATAAACCT GGTCACCTGA AATTGAAATT GAGCCACTTC CTTGAARAAT    414

CAAAATTCCT GTTAATAAAA RAAAAACAAA TGTAATTGAA ATAGCACACA GCATTCTCTA    474

GTCAATATCT TTAGTGATCT TCTTTAATAA ACATGAAAGC AAAAAAAAAA AA            526
```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo Sapiens

(ix) FEATURE:
- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 1..17
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.2
  seq LLLITAILAVAVG/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
Met Lys Lys Val Leu Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val
  1               5                  10                  15

Gly
```

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 822 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(D) DEVELOPMENTAL STAGE: Fetal
(F) TISSUE TYPE: kidney

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 260..464
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 153..357
id H57434
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 118..184
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 98..164
id H57434
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 56..113
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 35..92
id H57434
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 454..485
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 348..379
id H57434
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 118..545
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 1..428
id N27248
est

(ix) FEATURE:
(A) NAME/KEY: other

```
        (B) LOCATION: 65..369
        (C) IDENTIFICATION METHOD: blastn
        (D) OTHER INFORMATION:  identity 98
                                region 41..345
                                id H94779
                                est

    (ix) FEATURE:
        (A) NAME/KEY: other
        (B) LOCATION: 61..399
        (C) IDENTIFICATION METHOD: blastn
        (D) OTHER INFORMATION:  identity 99
                                region 6..344
                                id H09880
                                est

    (ix) FEATURE:
        (A) NAME/KEY: other
        (B) LOCATION: 408..458
        (C) IDENTIFICATION METHOD: blastn
        (D) OTHER INFORMATION:  identity 92
                                region 355..405
                                id H09880
                                est

    (ix) FEATURE:
        (A) NAME/KEY: other
        (B) LOCATION: 60..399
        (C) IDENTIFICATION METHOD: blastn
        (D) OTHER INFORMATION:  identity 97
                                region 56..395
                                id H29351
                                est

    (ix) FEATURE:
        (A) NAME/KEY: other
        (B) LOCATION: 393..432
        (C) IDENTIFICATION METHOD: blastn
        (D) OTHER INFORMATION:  identity 90
                                region 391..430
                                id H29351
                                est

    (ix) FEATURE:
        (A) NAME/KEY: sig_peptide
        (B) LOCATION: 346..408
        (C) IDENTIFICATION METHOD: Von Heijne matrix
        (D) OTHER INFORMATION:  score 5.5
                                seq SFLPSALVIWTSA/AF

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ACTCCTTTTA GCATAGGGGC TTCGGCGCCA GCGGCCAGCG CTAGTCGGTC TGGTAAGTGC     60

CTGATGCCGA GTTCCGTCTC TCGCGTCTTT TCCTGGTCCC AGGCAAAGCG GASGNAGATC    120

CTCAAACGGC CTAGTGCTTC GCGCTTCCGG AGAAAATCAG CGGTCTAATT AATTCCTCTG    180

GTTTGTTGAA GCAGTTACCA AGAATCTTCA ACCCTTTCCC ACAAAAGCTA ATTGAGTACA    240
```

```
CGTTCCTGTT GAGTACACGT TCCTGTTGAT TTACAAAAGG TGCAGGTATG AGCAGGTCTG    300

AAGACTAACA TTTTGTGAAG TTGTAAAACA GAAAACCTGT TAGAA ATG TGG TGG TTT    357
                                                  Met Trp Trp Phe
                                                      -20

CAG CAA GGC CTC AGT TTC CTT CCT TCA GCC CTT GTA ATT TGG ACA TCT      405
Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser
        -15                 -10                     -5

GCT GCT TTC ATA TTT TCA TAC ATT ACT GCA GTA ACA CTC CAC CAT ATA      453
Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile
    1               5                   10                  15

GAC CCG GCT TTA CCT TAT ATC AGT GAC ACT GGT ACA GTA GCT CCA RAA      501
Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa
                20                  25                  30

AAA TGC TTA TTT GGG GCA ATG CTA AAT ATT GCG GCA GTT TTA TGT CAA      549
Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln
            35                  40                  45

AAA TAGAAATCAG GAARATAATT CAACTTAAAG AAKTTCATTT CATGACCAAA           602
Lys

CTCTTCARAA ACATGTCTTT ACAAGCATAT CTCTTGTATT GCTTTCTACA CTGTTGAATT    662

GTCTGGCAAT ATTTCTGCAG TGGAAAATTT GATTTARMTA GTTCTTGACT GATAAATATG    722

GTAAGGTGGG CTTTTCCCCC TGTGTAATTG GCTACTATGT CTTACTGAGC CAAGTTGTAW    782

TTTGAAATAA AATGATATGA GAGTGACACA AAAAAAAAAA                           822
```

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..21
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.5
seq SFLPSALVIWTSA/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```
Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val
 1               5                  10                  15
```

```
Ile Trp Thr Ser Ala
            20


(2) INFORMATION FOR SEQ ID NO: 21:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 405 base pairs
            (B) TYPE: NUCLEIC ACID
            (C) STRANDEDNESS: DOUBLE
            (D) TOPOLOGY: LINEAR

      (ii) MOLECULE TYPE: CDNA

      (vi) ORIGINAL SOURCE:
            (A) ORGANISM: Homo Sapiens
            (F) TISSUE TYPE: Testis

      (ix) FEATURE:
            (A) NAME/KEY: other
            (B) LOCATION: complement(103..398)
            (C) IDENTIFICATION METHOD: blastn
            (D) OTHER INFORMATION:  identity 96
                                    region 1..296
                                    id AA442893
                                    est

      (ix) FEATURE:
            (A) NAME/KEY: sig_peptide
            (B) LOCATION: 185..295
            (C) IDENTIFICATION METHOD: Von Heijne matrix
            (D) OTHER INFORMATION:  score 5.9
                                    seq LSYASSALSPCLT/AP

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATCACCTTCT TCTCCATCCT TSTCTGGGCC AGTCCCCARC CCAGTCCCTC TCCTGACCTG      60

CCCAGCCCAA GTCAGCCTTC AGCACGCGCT TTTCTGCACA CAGATATTCC AGGCCTACCT     120

GGCATTCCAG GACCTCCGMA ATGATGCTCC AGTCCCTTAC AAGCGCTTCC TGGATGAGGG     180

TGGC ATG GTG CTG ACC ACC CTC CCC TTG CCC TCT GCC AAC AGC CCT GTG      229
     Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val
         -35                 -30                 -25

AAC ATG CCC ACC ACT GGC CCC AAC AGC CTG AGT TAT GCT AGC TCT GCC       277
Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala
        -20                 -15                 -10

CTG TCC CCC TGT CTG ACC GCT CCA AAK TCC CCC CGG CTT GCT ATG ATG       325
Leu Ser Pro Cys Leu Thr Ala Pro Xaa Ser Pro Arg Leu Ala Met Met
     -5                   1               5                  10

CCT GAC AAC TAAATATCCT TATCCAAATC AATAAARWRA RAATCCTCCC TCCARAAGGG    384
Pro Asp Asn


TTTCTAAAAA CAAAAAAAAA A                                                405
```

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..37
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.9
seq LSYASSALSPCLT/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```
Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn
  1               5                  10                  15

Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu
            20                  25                  30

Ser Pro Cys Leu Thr
        35
```

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 496 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 149..331
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 1..183
id AA397994
est

(ix) FEATURE:
(A) NAME/KEY: other

```
          (B) LOCATION: 328..485
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION:  identity 96
                                  region 179..336
                                  id AA397994
                                  est

     (ix) FEATURE:
          (A) NAME/KEY: other
          (B) LOCATION: complement(182..496)
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION:  identity 97
                                  region 14..328
                                  id AA399680
                                  est

     (ix) FEATURE:
          (A) NAME/KEY: sig_peptide
          (B) LOCATION: 196..240
          (C) IDENTIFICATION METHOD: Von Heijne matrix
          (D) OTHER INFORMATION:  score 5.5
                                  seq ILSTVTALTFAXA/LD

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

AAAAAATTGG TCCCAGTTTT CACCCTGCCG CAGGGCTGGC TGGGGAGGGC AGCGGTTTAG      60

ATTAGCCGTG GCCTAGGCCG TTTAACGGGG TGACACGAGC NTGCAGGGCC GAGTCCAAGG     120

CCCGGAGATA GGACCAACCG TCAGGAATGC GAGGAATGTT TTTCTTCGGA CTCTATCGAG     180

GCACACAGAC AGACC ATG GGG ATT CTG TCT ACA GTG ACA GCC TTA ACA TTT     231
                 Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
                 -15                 -10                 -5

GCC ARA GCC CTG GAC GGC TGC AGA AAT GGC ATT GCC CAC CCT GCA AGT      279
Ala Xaa Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
            1               5                  10

GAG AAG CAC AGA CTC GAG AAA TGT AGG GAA CTC GAG ASC ASC CAC TCG      327
Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Xaa Xaa His Ser
    15                  20                  25

GCC CCA GGA TCA ACC CAS CAC CGA AGA AAA ACA ACC AGA AGA AAT TAT      375
Ala Pro Gly Ser Thr Xaa His Arg Arg Lys Thr Thr Arg Arg Asn Tyr
30                  35                  40                  45

TCT TCA GCC TGAAATGAAK CCGGGATCAA ATGGTTGCTG ATCARAGCCC ATATTTAAAT   434
Ser Ser Ala

TGGAAAAGTC AAATTGASCA TTATTAAATA AAGCTTGTTT AATATGTCTC AAACAAAAAA     494

AA                                                                    496


(2) INFORMATION FOR SEQ ID NO: 24:

     (i) SEQUENCE CHARACTERISTICS:
```

(A) LENGTH: 15 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..15
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.5
seq ILSTVTALTFAXA/LD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

```
Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Xaa Ala
  1             5                   10                  15
```

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 623 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Testis

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 49..96
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.1
seq LVLTLCTLPLAVA/SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```
AAAGATCCCT GCAGCCCGGC AGGAGAGAAG GCTGAGCCTT CTGGCGTC ATG GAG AGG       57
                                                     Met Glu Arg
                                                         -15

CTC GTC CTA ACC CTG TGC ACC CTC CCG CTG GCT GTG GCG TCT GCT GGC       105
Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
            -10                  -5                  1

TGC GCC ACG ACG CCA GCT CGC AAC CTG AGC TGC TAC CAG TGC TTC AAG       153
Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys
     5                   10                  15

GTC AGC AGC TGG ACG GAG TGC CCG CCC ACC TGG TGC AGC CCG CTG GAC       201
```

```
Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp
20                  25                  30                  35

CAA GTC TGC ATC TCC AAC GAG GTG GTC GTC TCT TTT AAA TGG AGT GTA      249
Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val
                40                  45                  50

CGC GTC CTG CTC AGC AAA CGC TGT GCT CCC AGA TGT CCC AAC GAC AAC      297
Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn
            55                  60                  65

ATG AAK TTC GAA TGG TCG CCG GCC CCC ATG GTG CAA GGC GTG ATC ACC      345
Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr
        70                  75                  80

AGG CGC TGC TGT TCC TGG GCT CTC TGC AAC AGG GCA CTG ACC CCA CAG      393
Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln
    85                  90                  95

GAG GGG CGC TGG GCC CTG CRA GGG GGG CTC CTG CTC CAG GAC CCT TCG      441
Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Leu Gln Asp Pro Ser
100                 105                 110                 115

AGG GGC ARA AAA ACC TGG GTG CGG CCA CAG CTG GGG CTC CCA CTC TGC      489
Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys
                120                 125                 130

CTT CCC AWT TCC AAC CCC CTC TGC CCA RGG GAA ACC CAG GAA GGA          534
Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly
            135                 140                 145

TAACACTGTG GGTGCCCCCA CCTGTGCATT GGGACCACRA CTTCACCCTC TTGGARACAA    594

TAAACTCTCA TGCCCCCAAA AAAAAAAAA                                      623
```

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..16
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.1
seq LVLTLCTLPLAVA/SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

```
Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala
  1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 848 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(D) DEVELOPMENTAL STAGE: Fetal
(F) TISSUE TYPE: kidney

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 32..73
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.7
seq LWLLFFLVTAIHA/EL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

```
AACTTTGCCT TGTGTTTTCC ACCCTGAAAG A ATG TTG TGG CTG CTC TTT TTT CTG      55
                                   Met Leu Trp Leu Leu Phe Phe Leu
                                                   -10

GTG ACT GCC ATT CAT GCT GAA CTC TGT CAA CCA GGT GCA GAA AAT GCT        103
Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn Ala
    -5                  1               5                  10

TTT AAA GTG AGA CTT AGT ATC AGA ACA GCT CTG GGA GAT AAA GCA TAT        151
Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala Tyr
                15                  20                  25

GCC TGG GAT ACC AAT GAA GAA TAC CTC TTC AAA GCG ATG GTA GCT TTC        199
Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala Phe
            30                  35                  40

TCC ATG AGA AAA GTT CCC AAC AGA GAA GCA ACA GAA ATT TCC CAT GTC        247
Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His Val
        45                  50                  55

CTA CTT TGC AAT GTA ACC CAG AGG GTA TCA TTC TGG TTT GTG GTT ACA        295
Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val Thr
    60                  65                  70

GAC CCT TCA AAA AAT CAC ACC CTT CCT GCT GTT GAG GTG CAA TCA GCC        343
Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser Ala
75                  80                  85                  90

ATA AGA ATG AAC AAG AAC CGG ATC AAC AAT GCC TTC TTT CTA AAT GAC        391
Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn Asp
                95                  100                 105

CAA ACT CTG GAA TTT TTA AAA ATC CCT TCC ACA CTT GCA CCA CCC ATG        439
```

```
Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro Met
            110                 115                 120

GAC CCA TCT GTG CCC ATC TGG ATT ATT ATA TTT GGT GTG ATA TTT TGC      487
Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe Cys
        125                 130                 135

ATC ATC ATA GTT GCA ATT GCA CTA CTG ATT TTA TCA GGG ATC TGG CAA      535
Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp Gln
    140                 145                 150

CGT ADA ARA AAG AAC AAA GAA CCA TCT GAA GTG GAT GAC GCT GAA RAT      583
Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu Xaa
155                 160                 165                 170

AAK TGT GAA AAC ATG ATC ACA ATT GAA AAT GGC ATC CCC TCT GAT CCC      631
Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp Pro
                175                 180                 185

CTG GAC ATG AAG GGA GGG CAT ATT AAT GAT GCC TTC ATG ACA GAG GAT      679
Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu Asp
            190                 195                 200

GAG AGG CTC ACC CCT CTC TGAAGGGCTG TTGTTCTGCT TCCTCAARAA             727
Glu Arg Leu Thr Pro Leu
        205

ATTAAACATT TGTTTCTGTG TGACTGCTGA GCATCCTGAA ATACCAAGAG CAGATCATAT    787

WTTTTGTTTC ACCATTCTTC TTTTGTAATA AATTTTGAAT GTGCTTGAAA AAAAAAAAAA    847

C                                                                    848
```

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: AMINO ACID
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..14
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.7
seq LWLLFFLVTAIHA/EL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

```
Met Leu Trp Leu Leu Phe Phe Leu Val Thr Ala Ile His Ala
 1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGGAAGATGG AGATAGTATT GCCTG 25

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTGCCATGTA CATGATAGAG AGATTC 26

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 546 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..517

(ix) FEATURE:
(A) NAME/KEY: transcription start site
(B) LOCATION: 518

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 17..25
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name CMYB_01
score 0.983
sequence TGTCAGTTG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(18..27)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MYOD_Q6
score 0.961
sequence CCCAACTGAC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(75..85)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name S8_01
score 0.960
sequence AATAGAATTAG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 94..104
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name S8_01
score 0.966
sequence AACTAAATTAG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(129..139)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name DELTAEF1_01
score 0.960
sequence GCACACCTCAG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(155..165)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA_C
score 0.964
sequence AGATAAATCCA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 170..178
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name CMYB_01
score 0.958
sequence CTTCAGTTG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 176..189
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA1_02
score 0.959
sequence TTGTAGATAGGACA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 180..190
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA_C

score 0.953
sequence AGATAGGACAT

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 284..299
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name TAL1ALPHAE47_01
score 0.973
sequence CATAACAGATGGTAAG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 284..299
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name TAL1BETAE47_01
score 0.983
sequence CATAACAGATGGTAAG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 284..299
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name TAL1BETAITF2_01
score 0.978
sequence CATAACAGATGGTAAG

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(287..296)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MYOD_Q6
score 0.954
sequence ACCATCTGTT

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(302..314)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA1_04
score 0.953
sequence TCAAGATAAAGTA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 393..405
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name IK1_01
score 0.963
sequence AGTTGGGAATTCC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 393..404
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name IK2_01
score 0.985
sequence AGTTGGGAATTC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site

(B) LOCATION: 396..405
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name CREL_01
score 0.962
sequence TGGGAATTCC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 423..436
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA1_02
score 0.950
sequence TCAGTGATATGGCA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(478..489)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name SRY_02
score 0.951
sequence TAAAACAAAACA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 486..493
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name E2F_02
score 0.957
sequence TTTAGCGC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(514..521)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.975
sequence TGAGGGGA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TGAGTGGAGT GTTACATGTC AGTTGGGTTA AGTTTGTTAA TGTCATTCAA ATCTTCTATG 60

TCTTGATTTG CCTGCTAATT CTATTATTTC TGGAACTAAA TTAGTTTGAT GGTTCTATTA 120

GTTATTGACT GAGGTGTGCT AATCTCCCAT TATGTGGATT TATCTATTTC TTCAGTTGTA 180

GATAGGACAT TGATAGATAC ATAAGTACCA GGACAAAAGC AGGGAGATCT TTTTTCCAAA 240

ATCAGGAGAA AAAAATGACA TCTGGAAAAC CTATAGGGAA AGGCATAACA GATGGTAAGG 300

ATACTTTATC TTGAGTAGGA GAGCCTTCCT GTGGCAACGT GGAGAAGGGA AGAGGTCGTA 360

GAATTGAGGA GTCAGCTCAG TTAGAAGCAG GGAGTTGGGA ATTCCGTTCA TGTGATTTAG 420

CATCAGTGAT ATGGCAAATG TGGGACTAAG GGTAGTGATC AGAGGGTTAA AATTGTGTGT 480

TTTGTTTTAG CGCTGCTGGG GCATCGCCTT GGGTCCCCTC AAACAGATTC CCATGAATCT 540

CTTCAT 546

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GTACCAGGGA CTGTGACCAT TGC 23

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CTGTGACCAT TGCTCCCAAG AGAG 24

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 861 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..806

(ix) FEATURE:
(A) NAME/KEY: transcription start site
(B) LOCATION: 807

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(60..70)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name NFY_Q6
score 0.956
sequence GGACCAATCAT

```
(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 70..77
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name MZF1_01
                              score 0.962
                              sequence CCTGGGGA

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 124..132
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name CMYB_01
                              score 0.994
                              sequence TGACCGTTG

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(126..134)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name VMYB_02
                              score 0.985
                              sequence TCCAACGGT

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 135..143
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name STAT_01
                              score 0.968
                              sequence TTCCTGGAA

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(135..143)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name STAT_01
                              score 0.951
                              sequence TTCCAGGAA

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(252..259)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name MZF1_01
                              score 0.956
                              sequence TTGGGGGA

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 357..368
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name IK2_01
                              score 0.965
                              sequence GAATGGGATTTC

(ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 384..391
      (C) IDENTIFICATION METHOD: matinspector prediction
```

(D) OTHER INFORMATION: name MZF1_01
score 0.986
sequence AGAGGGGA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(410..421)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name SRY_02
score 0.955
sequence GAAAACAAAACA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 592..599
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.960
sequence GAAGGGGA

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 618..627
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MYOD_Q6
score 0.981
sequence AGCATCTGCC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: 632..642
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name DELTAEF1_01
score 0.958
sequence TCCCACCTTCC

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(813..823)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name S8_01
score 0.992
sequence GAGGCAATTAT

(ix) FEATURE:
(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(824..831)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.986
sequence AGAGGGGA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TACTATAGGG CACGCGTGGT CGACGGCCGG GCTGTTCTGG AGCAGAGGGC ATGTCAGTAA 60

TGATTGGTCC CTGGGGAAGG TCTGGCTGGC TCCAGCACAG TGAGGCATTT AGGTATCTCT 120

CGGTGACCGT TGGATTCCTG GAAGCAGTAG CTGTTCTGTT TGGATCTGGT AGGGACAGGG 180

CTCAGAGGGC TAGGCACGAG GGAAGGTCAG AGGAGAAGGS AGGSARGGCC CAGTGAGARG 240

GGAGCATGCC TTCCCCCAAC CCTGGCTTSC YCTTGGYMAM AGGGCGKTTY TGGGMACTTR 300

AAYTCAGGGC CCAASCAGAA SCACAGGCCC AKTCNTGGCT SMAAGCACAA TAGCCTGAAT 360

GGGATTTCAG GTTAGNCAGG GTGAGAGGGG AGGCTCTCTG GCTTAGTTTT GTTTTGTTTT 420

CCAAATCAAG GTAACTTGCT CCCTTCTGCT ACGGGCCTTG GTCTTGGCTT GTCCTCACCC 480

AGTCGGAACT CCCTACCACT TTCAGGAGAG TGGTTTTAGG CCCGTGGGGC TGTTCTGTTC 540

CAAGCAGTGT GAGAACATGG CTGGTAGAGG CTCTAGCTGT GTGCGGGGCC TGAAGGGGAG 600

TGGGTTCTCG CCCAAAGAGC ATCTGCCCAT TTCCCACCTT CCCTTCTCCC ACCAGAAGCT 660

TGCCTGAGCT GTTTGGACAA AAATCCAAAC CCCACTTGGC TACTCTGGCC TGGCTTCAGC 720

TTGGAACCCA ATACCTAGGC TTACAGGCCA TCCTGAGCCA GGGGCCTCTG GAAATTCTCT 780

TCCTGATGGT CCTTTAGGTT TGGGCACAAA ATATAATTGC CTCTCCCCTC TCCCATTTTC 840

TCTCTTGGGA GCAATGGTCA C 861

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTGGGATGGA AGGCACGGTA 20

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GAGACCACAC AGCTAGACAA 20

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 555 base pairs
  (B) TYPE: NUCLEIC ACID
  (C) STRANDEDNESS: DOUBLE
  (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:
  (A) NAME/KEY: promoter
  (B) LOCATION: 1..500

(ix) FEATURE:
  (A) NAME/KEY: transcription start site
  (B) LOCATION: 501

(ix) FEATURE:
  (A) NAME/KEY: TF binding-site
  (B) LOCATION: 191..206
  (C) IDENTIFICATION METHOD: matinspector prediction
  (D) OTHER INFORMATION: name ARNT_01
    score 0.964
    sequence GGACTCACGTGCTGCT

(ix) FEATURE:
  (A) NAME/KEY: TF binding-site
  (B) LOCATION: 193..204
  (C) IDENTIFICATION METHOD: matinspector prediction
  (D) OTHER INFORMATION: name NMYC_01
    score 0.965
    sequence ACTCACGTGCTG

(ix) FEATURE:
  (A) NAME/KEY: TF binding-site
  (B) LOCATION: 193..204
  (C) IDENTIFICATION METHOD: matinspector prediction
  (D) OTHER INFORMATION: name USF_01
    score 0.985
    sequence ACTCACGTGCTG

(ix) FEATURE:
  (A) NAME/KEY: TF binding-site
  (B) LOCATION: complement(193..204)
  (C) IDENTIFICATION METHOD: matinspector prediction
  (D) OTHER INFORMATION: name USF_01
    score 0.985
    sequence CAGCACGTGAGT

(ix) FEATURE:
  (A) NAME/KEY: TF binding-site
  (B) LOCATION: complement(193..204)
  (C) IDENTIFICATION METHOD: matinspector prediction
  (D) OTHER INFORMATION: name NMYC_01
    score 0.956
    sequence CAGCACGTGAGT

(ix) FEATURE:
  (A) NAME/KEY: TF binding-site
  (B) LOCATION: complement(193..204)
  (C) IDENTIFICATION METHOD: matinspector prediction

```
      (D) OTHER INFORMATION:  name MYCMAX_02
                              score 0.972
                              sequence CAGCACGTGAGT

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 195..202
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name USF_C
                              score 0.997
                              sequence TCACGTGC

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(195..202)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name USF_C
                              score 0.991
                              sequence GCACGTGA

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(210..217)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name MZF1_01
                              score 0.968
                              sequence CATGGGGA

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 397..410
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name ELK1_02
                              score 0.963
                              sequence CTCTCCGGAAGCCT

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: 400..409
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name CETS1P54_01
                              score 0.974
                              sequence TCCGGAAGCC

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(460..470)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name AP1_Q4
                              score 0.963
                              sequence AGTGACTGAAC

 (ix) FEATURE:
      (A) NAME/KEY: TF binding-site
      (B) LOCATION: complement(460..470)
      (C) IDENTIFICATION METHOD: matinspector prediction
      (D) OTHER INFORMATION:  name AP1FJ_Q2
                              score 0.961
                              sequence AGTGACTGAAC

 (ix) FEATURE:
```

(A) NAME/KEY: TF binding-site
(B) LOCATION: 547..555
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name PADS_C
score 1.000
sequence TGTGGTCTC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

```
CTATAGGGCA CGCKTGGTCG ACGGCCCGGG CTGGTCTGGT CTGTKGTGGA GTCGGGTTGA    60
AGGACAGCAT TTGTKACATC TGGTCTACTG CACCTTCCCT CTGCCGTGCA CTTGGCCTTT   120
KAWAAGCTCA GCACCGGTGC CCATCACAGG GCCGGCAGCA CACACATCCC ATTACTCAGA   180
AGGAACTGAC GGACTCACGT GCTGCTCCGT CCCCATGAGC TCAGTGGACC TGTCTATGTA   240
GAGCAGTCAG ACAGTGCCTG GGATAGAGTG AGAGTTCAGC CAGTAAATCC AAGTGATTGT   300
CATTCCTGTC TGCATTAGTA ACTCCCAACC TAGATGTGAA AACTTAGTTC TTTCTCATAG   360
GTTGCTCTGC CCATGGTCCC ACTGCAGACC CAGGCACTCT CCGGAAGCCT GGAAATCACC   420
CGTGTCTTCT GCCTGCTCCC GCTCACATCC CACACTTGTG TTCAGTCACT GAGTTACAGA   480
TTTTGCCTCC TCAATTTCTC TTGTCTTAGT CCCATCCTCT GTTCCCCTGG CCAGTTTGTC   540
TAGCTGTGTG GTCTC                                                     555
```

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 120 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 16..84
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 11.4
seq VLALLLFVHYSNG/DE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

```
ACTTCCTGGT GCTGC ATG GTG TTC GTG CAC CTG TAC CTG GGT AAC GTG CTG    51
                 Met Val Phe Val His Leu Tyr Leu Gly Asn Val Leu
                             -20                 -15

GCG CTG CTG CTC TTC GTG CAC TAC AGC AAC GGC GAC GAA AGC AGC GAT      99
```

```
Ala Leu Leu Leu Phe Val His Tyr Ser Asn Gly Asp Glu Ser Ser Asp
    -10                 -5                   1               5

CCC GGG CCC CAR CAC CGT GCC                                       120
Pro Gly Pro Gln His Arg Ala
                10
```

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 202..288
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 11.3
seq FLLCIFLICAALA/AQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

```
AAAAGTGGAA AATGGGAGGC ATGAAATACA TCTTTTCGTT GTTGTTCTTT CTTTTGCTAG   60

AAGGAGGCAA AACAGAGCAA GTAAAACATT CAGAGACATA TTGCATGTTT CAAGACAAGA  120

AGTACAGAGT GGGTGAGAGA TGGCATCCTT ACCTGGAACC TTATGGGTTG GTTTACTGCG  180

TGAACTGCAT CTGCTCAGAG A ATG GGA ATG TGC TTT GCA GCC GAG TCA GAT    231
                        Met Gly Met Cys Phe Ala Ala Glu Ser Asp
                                            -25                 -20

GTC CAA ATG TTC ATT GCC TTT CTC CTG TGC ATA TTC CTC ATC TGT GCT    279
Val Gln Met Phe Ile Ala Phe Leu Leu Cys Ile Phe Leu Ile Cys Ala
                -15                 -10                     -5

GCC CTC GCT GCC CAG AAG AGT GGG                                    303
Ala Leu Ala Ala Gln Lys Ser Gly
            1               5
```

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 313 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 203..280
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 11
seq VLFLFLFWGVSLA/GS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

```
AAGGATGCTA TGCAAGTCAC TAATAAAGGA AGACACGGAC AGATGAACTT AAAAGAGAAG    60

CTTTAGCTGC CAAAGATTGG GAAAGGGAAA GGMCAAAAAA GACCCCTGGG CTACACGGCG   120

TAGGTGCAGG GTTTCCTACT GCTGTTCTTT TATGCTGGGA GCTGTGGCTG TAACCAACTA   180

GGAAATAACG TATGCAGCAG CT ATG GCT GTC AGA GAG TTG TGC TTC TCA AGA    232
                         Met Ala Val Arg Glu Leu Cys Phe Ser Arg
                             -25                 -20

CAA AGG CAA GTC CTG TTT CTT TTT CTT TTT TGG GGA GTG TCC TTG GCA     280
Gln Arg Gln Val Leu Phe Leu Phe Leu Phe Trp Gly Val Ser Leu Ala
    -15                 -10                 -5

GGT TCT GGG TTT GGA CGT TAT TCG GTG ACC GGG                         313
Gly Ser Gly Phe Gly Arg Tyr Ser Val Thr Gly
  1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 323 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 117..170
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.7
seq LILLALATGLVGG/ET

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

```
AAAACCCAGC CTACCTGCTG TAGCTGCCGC CACTGCCGTC TCCGCCGCCA CTGGWCCCCC    60
```

```
AGAGCBNMAG CCCCAGAGCC TAGGAACCTG GGGCCCGCTC CTCCCCCCTC CAGGCC ATG  119
                                                              Met

AGG ATT CTG CAG TTA ATC CTG CTT GCT CTG GCA ACA GGG CTT GTA GGG   167
Arg Ile Leu Gln Leu Ile Leu Leu Ala Leu Ala Thr Gly Leu Val Gly
        -15                 -10                 -5

GGA GAG ACC AGG ATC ATC AAG GGG TTC GAG TGC AAG CCT CAC TCC CAG   215
Gly Glu Thr Arg Ile Ile Lys Gly Phe Glu Cys Lys Pro His Ser Gln
      1               5                  10                  15

CCC TGG CAG GCA GCC CTG TTC GAG AAG ACG CGG CTA CTC TGT GGG GCG   263
Pro Trp Gln Ala Ala Leu Phe Glu Lys Thr Arg Leu Leu Cys Gly Ala
                20                  25                  30

ACG CTC ATC GCC CCC AGA TGG CTC CTG ACA GCA GCC CAC TGC CTC AAG   311
Thr Leu Ile Ala Pro Arg Trp Leu Leu Thr Ala Ala His Cys Leu Lys
            35                  40                  45

CCC CGC TAC GGG                                                   323
Pro Arg Tyr Gly
        50
```

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 264 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 94..147
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.7
seq LILLALATGLVGG/ET

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

```
AAGAGGTTGA GGTGGCTGCG GGACTGGAAG TCATCGGGCA GAGGTCTCAC AGCAGCCAAG   60

AAACCTGGGG CCCGCTCCTC CCCCCTCCAG GCC ATG AGG ATT CTG CAG TTA ATC  114
                                     Met Arg Ile Leu Gln Leu Ile
                                                 -15

CTG CTT GCT CTG GCA ACA GGG CTT GTA GGG GGA GAG ACC AGG ATC ATC   162
Leu Leu Ala Leu Ala Thr Gly Leu Val Gly Gly Glu Thr Arg Ile Ile
    -10                 -5                  1               5

AAG GGG TTC GAG TGC AAG CCT CAC TNC CAG CCC TGG CAG GCA GCC CTG   210
```

```
Lys Gly Phe Glu Cys Lys Pro His Xaa Gln Pro Trp Gln Ala Ala Leu
                10                  15                  20

TTC GAG AAG ACG CGG CTA CTC TGT GGG GCG ACG CTC ATC GCC CCC AGA      258
Phe Glu Lys Thr Arg Leu Leu Cys Gly Ala Thr Leu Ile Ala Pro Arg
            25                  30                  35

TGG CTC
Trp Leu                                                              264
```

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 331 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 23..112
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.6
seq SLLLAVLVFFLFA/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

```
CTCTAGAACC CGACCCACCA CC ATG AGG TCC TGC CTG TGG AGA TGC AGG CAC      52
                         Met Arg Ser Cys Leu Trp Arg Cys Arg His
                         -30                 -25

CTG AGC CAA GGC GTC CAG TGG TCC TTG CTT CTG GCT GTC CTG GTC TTC      100
Leu Ser Gln Gly Val Gln Trp Ser Leu Leu Leu Ala Val Leu Val Phe
-20                 -15                 -10                 -5

TTT CTC TTC GCC TTG CCC TCT DNH AVT TRR KGD SCT CAA ACA AAG CCT      148
Phe Leu Phe Ala Leu Pro Ser Xaa Xaa Xaa Xaa Xaa Gln Thr Lys Pro
                1                   5                   10

TCC AGG CAT CAA CGC ACA GAG AAC ATT AAA GAA AGG TCT CTA CWG TCC      196
Ser Arg His Gln Arg Thr Glu Asn Ile Lys Glu Arg Ser Leu Xaa Ser
        15                  20                  25

CTG GCA AAG CCT AAG TCC CAG GCA CCC ACA AGG GCA AGG AGG ACA ACC      244
Leu Ala Lys Pro Lys Ser Gln Ala Pro Thr Arg Ala Arg Arg Thr Thr
    30                  35                  40

ATC TAT GCA GAG CCA GTG CCA GAG AAC AAT GCC CTC AAC ACA CAA ACC      292
Ile Tyr Ala Glu Pro Val Pro Glu Asn Asn Ala Leu Asn Thr Gln Thr
45                  50                  55                  60

CAG CCC AAG GCC CAC ACC ACC GGA GAC AGA AGG AAA GGA                  331
```

```
Gln Pro Lys Ala His Thr Thr Gly Asp Arg Arg Lys Gly
                65                  70
```

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 406 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 167..220
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.6
seq XILLALATGLVGG/EI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

```
AATGTGGGAC GTGGCTTTGT TCTAATAAGA CGAAGGGTGG AGTGCAGGCT TGGAAAGCAG    60

GAGAGCTCAG CCTACGTCTT TAATCCTCCT GCCCACCCCT TGGRTTCTGT CTCCACTGGG   120

RCTCAAGASV AGGACCCTGG GGGCCCGCTC CTCCCCCCTC CAGGCC ATG AGG ATT      175
                                                   Met Arg Ile

CTG CAG TKA ATC CTG CTT GCT CTG GCA ACA GGG CTT GTA GGG GGA GAG     223
Leu Gln Xaa Ile Leu Leu Ala Leu Ala Thr Gly Leu Val Gly Gly Glu
-15                 -10                 -5                   1

ATC AGG ATC ATC AAG GGG TTC GAG TGC AAG CCT CAC TCC CAG CCC TGG     271
Ile Arg Ile Ile Lys Gly Phe Glu Cys Lys Pro His Ser Gln Pro Trp
            5                   10                  15

CAG GCA GCC CTG TTC GAG AAG ACG CGG CTA CTA CTG TGG GGC GAC GCT     319
Gln Ala Ala Leu Phe Glu Lys Thr Arg Leu Leu Leu Trp Gly Asp Ala
        20                  25                  30

CAT CGC CCC CAG ATG GCT CCT GAC AGC AGC CCA CTG CCT CAA GCC CCG     367
His Arg Pro Gln Met Ala Pro Asp Ser Ser Pro Leu Pro Gln Ala Pro
    35                  40                  45

CTA CAT AGT TCA CCT GGG GCA GCA CAA CCT CCA GAA GGA                 406
Leu His Ser Ser Pro Gly Ala Ala Gln Pro Pro Glu Gly
 50                 55                  60
```

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 187 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 35..148
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.4
seq LWLLLKLVSTXWA/VR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```
ATACTGTTTA TAAGCAACCT TGGTTTTACA TAGT ATG TTG GAA GAG TGT GGG GCT    55
                                      Met Leu Glu Glu Cys Gly Ala
                                                  -35

GGG GTT GAT TTA GGA TTT GGA GGT GTA AAG TTT GCC AGT GAG ACA CCA     103
Gly Val Asp Leu Gly Phe Gly Gly Val Lys Phe Ala Ser Glu Thr Pro
    -30                 -25                 -20

AAC CTT CTC TGG CTG CTT TTA AAA CTK GTA AGT ACC YCT TGG GCT GTA     151
Asn Leu Leu Trp Leu Leu Leu Lys Leu Val Ser Thr Xaa Trp Ala Val
-15                 -10                  -5                   1

AGA GTG ACT TTG ATC ATA TTT AAC AAC CAG GCA AGG                     187
Arg Val Thr Leu Ile Ile Phe Asn Asn Gln Ala Arg
            5                   10
```

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 329 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 249..317
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 10.2
seq RCLLLALVAESSS/QT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

```
ATCTACTATA AAATCGATAG AAAAAAAAGT TCTTTATGGC TACTGGTCAG CTTTTATTCC   60

TGATACGCCT GAACTTGGCA GCCCACAGTC AGTGTCCTTG ATGACTCTTA SATTGAAAGA  120

CCCKTCTTCC AAAGACACGT GCCTGTGCTC TGCAAGTTTK ATCTGCCATC TTGGAAGGCT  180

CAAAGCAGTT TCTTTCTGTT GCTGAAGATA CCAGTGACCA CAGAAGGGCT TTTACCCCCT  240

TCTCCGTA ATG ATC GCT TGC AGC ATT AGA GAG TTG CAC AGA TGT CTK TTG   290
         Met Ile Ala Cys Ser Ile Arg Glu Leu His Arg Cys Leu Leu
                     -20                 -15                 -10

TTA GCT TTG GTG GCG GAG TCA TCC TCA CAG ACC CAC GGG                329
Leu Ala Leu Val Ala Glu Ser Ser Ser Gln Thr His Gly
                -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 277 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 182..232
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.2
  seq SLVLCLLSATVFS/LQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

```
AGTTTTTTCC AGCTCCTGGG CGAATCCCAC ATCTGTTTCA ACTCTCCGCC GAGGGCGAGC   60

AGGAGCGAGA GTGTGTCGAA TCTGCGAGTG AAGAGGGAAC SAGGGGAAAA GAAACAAAGC  120

CACAGACGCA ACTTGAGACT CCCGCATCCC AAAAGAAGCA CCAGATCAGC AAAAAAAGAA  180

G ATG GGC CCC CCG AGC CTC GTG CTG TGC TTG CTG TCC GCA ACT GTG TTC  229
  Met Gly Pro Pro Ser Leu Val Leu Cys Leu Leu Ser Ala Thr Val Phe
          -15                 -10                  -5

TCC CTG CAG GGT GGA AGC TCG GCC TTC CTG TCG CAC CAC CGC CCC GGG    277
Ser Leu Gln Gly Gly Ser Ser Ala Phe Leu Ser His His Arg Pro Gly
      1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 17..121
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 9
seq AMWWLLLWGVLQX/XP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

```
AGATGTCCAG TTCCAG ATG CCT GGA CCC AGA GTG TGG GGG AAA TAT CTC TGG    52
                  Met Pro Gly Pro Arg Val Trp Gly Lys Tyr Leu Trp
                  -35                 -30                 -25

AGA AGC CCT CAC TCC AAA GGC TGT CCA GGC GCA ATG TGG TGG CTG CTT     100
Arg Ser Pro His Ser Lys Gly Cys Pro Gly Ala Met Trp Trp Leu Leu
            -20                 -15                 -10

CTC TGG GGA GTC CTC CAG GST TKG CCC AAC CCG GGG CTC CGT CCT CTT     148
Leu Trp Gly Val Leu Gln Xaa Xaa Pro Asn Pro Gly Leu Arg Pro Leu
        -5                   1               5

GGC CHA AGA GCT ACC CCA GCA GCT GAC ATC CCC CGG GTA CCC AGA GCC     196
Gly Xaa Arg Ala Thr Pro Ala Ala Asp Ile Pro Arg Val Pro Arg Ala
10                  15                  20                  25

GTA TGG CAA AGG CCA AGA GAG CAR CAC GGA CAT CAA GGC TCC AGA GGG     244
Val Trp Gln Arg Pro Arg Glu Gln His Gly His Gln Gly Ser Arg Gly
                30                  35                  40

CTT TGC TGT GAG GCT CGT CTT CCA GGA CTT CGA CCT GGA GCC GTC CCA     292
Leu Cys Cys Glu Ala Arg Leu Pro Gly Leu Arg Pro Gly Ala Val Pro
            45                  50                  55

GGA CTG TGC AGG GGA CTC TGT CAC AAT CTC ATT CGT CGG TTC GGA TCC     340
Gly Leu Cys Arg Gly Leu Cys His Asn Leu Ile Arg Arg Phe Gly Ser
        60                  65                  70

AAG CCA CTC GGG                                                     352
Lys Pro Leu Gly
    75
```

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 base pairs

(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 151..216
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.8
seq LLTLALLGGPTWX/XK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

```
AAGAGCCCCA CGGCCAGCTC CTTCCTGTTC CCCTGGCGGC CCCTCGCTTC TTCCTTCTGG   60

ATGGGGGCCC AGGGGGCCAG GAGAGTATAA ASGSGWKDKG GARGGGTGCC CGGCACAACC  120

AGACGCCCAG TCACAGGCGA GAGCCCTGGG ATG CAC CGG CCA GAG GCC ATG CTG   174
                                 Met His Arg Pro Glu Ala Met Leu
                                         -20                 -15

CTG CTG CTC ACG CTT GCC CTC CTG GGG GGC CCC ACC TGG GMA SGG AAG    222
Leu Leu Leu Thr Leu Ala Leu Leu Gly Gly Pro Thr Trp Xaa Xaa Lys
                -10                 -5                      1

ATG TAT GGC CCT GGA GGA GGC AAG TAT TTC AGC ACC ACT GAA GAC TAC    270
Met Tyr Gly Pro Gly Gly Gly Lys Tyr Phe Ser Thr Thr Glu Asp Tyr
          5                  10                  15

GAC CAT GAA ATC ACA GGG CTG CGG GTG TCT GTA GGT CKT CTC CTG GTG    318
Asp His Glu Ile Thr Gly Leu Arg Val Ser Val Gly Xaa Leu Leu Val
     20                  25                  30

AAA AGT GTC CAG GTG AAA CTT GGA GAC TCC TGG GAC GTG AAA CTG GGA    366
Lys Ser Val Gln Val Lys Leu Gly Asp Ser Trp Asp Val Lys Leu Gly
 35                  40                  45                  50

GGC CTT AGG TGG GAA TAC CCA GGA AGT CAC CCT GCA GCC AGG CGA ATA    414
Gly Leu Arg Trp Glu Tyr Pro Gly Ser His Pro Ala Ala Arg Arg Ile
                 55                  60                  65

CAT CAC AAA AGT CTT TGT CGC TTC CAA GCT TTC CTC                    450
His His Lys Ser Leu Cys Arg Phe Gln Ala Phe Leu
             70                  75
```

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 181 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 5..49
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.6
seq SVSLALLSGWVGS/RQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

```
AGAC ATG GTA AGT GTG AGT TTA GCG CTG CTG TCC GGA TGG GTT GGT AGC      49
     Met Val Ser Val Ser Leu Ala Leu Leu Ser Gly Trp Val Gly Ser
     -15                 -10                 -5

AGA CAG GGT GGA GTA GGG TTA AGC ACA CTG GTC ACC TTA GGA TTG GTT       97
Arg Gln Gly Gly Val Gly Leu Ser Thr Leu Val Thr Leu Gly Leu Val
  1               5                  10                  15

TCC TGG TGC TGG AGA ATG GTT AGG ACA CAG GCC TTG GAA GGT TTT TTG      145
Ser Trp Cys Trp Arg Met Val Arg Thr Gln Ala Leu Glu Gly Phe Leu
            20                  25                  30

AGT GTG AAA TAT TAC TCA GCG TTT TCT GCA GAC CTG                      181
Ser Val Lys Tyr Tyr Ser Ala Phe Ser Ala Asp Leu
        35                  40
```

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 129..275
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.5
seq IVFLLLRVSPCLG/PS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

```
AAGATATCAC TGTCTTGTTT TCACTTAGAT CCTACTTACA AAGTGAGGGT TATTAACAGA     60
```

```
ATAAAGCCTT CCTTTAAAGC TTTATAATAA TCATATTTAT TAATAATGCT GTTGTGCATA  120

CTTATAGT ATG CAT ATA TTC AGC ATA TGT TGC ATG TST TCA GAA TTA CAT   170
         Met His Ile Phe Ser Ile Cys Cys Met Xaa Ser Glu Leu His
                         -45                 -40

AAG ATG AAA TCC CTT TCA TTG CAA CTT GCA AGT GAG AAA AGA TCC TTA   218
Lys Met Lys Ser Leu Ser Leu Gln Leu Ala Ser Glu Lys Arg Ser Leu
-35                 -30                 -25                 -20

GTG GCT CTG GTG GAA GAA ATA GTA TTT CTT CTT CTC AGG GTG TCT CCC   266
Val Ala Leu Val Glu Glu Ile Val Phe Leu Leu Leu Arg Val Ser Pro
            -15                 -10                     -5

TGC CTT GGC CCC TCC CAB AAG CCC CGG                               293
Cys Leu Gly Pro Ser Xaa Lys Pro Arg
             1               5
```

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 323 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 258..308
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.3
seq VSALLMAWFGVLS/CV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

```
AGCGCCGAGC TGACCGGGCG ACGCCGCGGG AGGTTCTGGA AACGCCGGGA GCTGCGAGTG   60

TCCAGACATC CTTGTGGAAC CAGGCGTTGT KTTTCCTTGG CAGCTGCGGA GACCCGTGAT  120

AATTCGTTAA CTAATTCAAC AAACGGGACC CTTCTGTGTG CCAGAAACCG CAAGCAGTTG  180

CTAACCCAGT GGGACAGGCG GATTGGAAGA GCGGGAAGGT CCTGGCCCAG AGCAGTGTGA  240

CACTTCCCTC TGTGACC ATG AAA CTC TGG GTG TCT GCA TTG CTG ATG GCC    290
                   Met Lys Leu Trp Val Ser Ala Leu Leu Met Ala
                           -15                 -10

TGG TTT GGT GTC CTG AGC TGT GTG CAG ACC GGG                       323
Trp Phe Gly Val Leu Ser Cys Val Gln Thr Gly
     -5                  1               5
```

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 235 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 92..157
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.3
seq LLLPLMLMSMVSS/SL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

```
AGACCTGAGT CATCCCCAGG GATCAGGAGC CTCCAGCAGG GAACCTTCCA TTATATTCTT    60

CAAGCAACTT ACAGCTGCAC CGACAGTTGC G ATG AAA GTT CTA ATC TCT TCC     112
                                   Met Lys Val Leu Ile Ser Ser
                                               -20

CTC CTC CTG TTG CTG CCA CTA ATG CTG ATG TCC ATG GTC TCT AGC AGC    160
Leu Leu Leu Leu Leu Pro Leu Met Leu Met Ser Met Val Ser Ser Ser
-15                 -10                  -5                   1

CTG AWT CCA GGG GTC GCC AGA GGC CAC AGG GAC CGA GGC CAG GCT TCT    208
Leu Xaa Pro Gly Val Ala Arg Gly His Arg Asp Arg Gly Gln Ala Ser
             5                  10                  15

AGG AGA TGG CTC CAG GAA GGC GGA CTG                                235
Arg Arg Trp Leu Gln Glu Gly Gly Leu
        20                  25
```

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide

(B) LOCATION: 159..224
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.3
seq LLLPLMLMSMVSS/SL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

```
ACTGTTCTCG CCCTCAAATG GGAACGCTGA CCTGGGACTA AAGCATAGAC CACCAGGCTG   60

AGTATCCTGA CCTGAGTCAT CCCCAGGGAT CAGGAGCCTC CAGCAGGGAA CCTTCCATTA  120

TATTCTTCAA GCAACTTACA GCTGCACCGA CAGTTGCG ATG AAA GTT CTA ATC TCT  176
                                          Met Lys Val Leu Ile Ser
                                                      -20

TCC CTC CTC CTG TTG CTG CCA CTA ATG CTG ATG TCC ATG GTC TCT AGC     224
Ser Leu Leu Leu Leu Leu Pro Leu Met Leu Met Ser Met Val Ser Ser
    -15                 -10                  -5

AGC CTG AAT CCA GGG GTC GCC AGA GGC CAC AGG GAC CGA GGC CAG GCT     272
Ser Leu Asn Pro Gly Val Ala Arg Gly His Arg Asp Arg Gly Gln Ala
  1             5                  10                  15

TCT AGG AGA TGG CTC CAG GAA GGC GGC CAA GAA TGT GAG TGC AAA GAT     320
Ser Arg Arg Trp Leu Gln Glu Gly Gly Gln Glu Cys Glu Cys Lys Asp
            20                  25                  30

TGG TTC CTG AGA GCC CCG AGA AGA AAA TTC ATG ACA GTG TCT GGG         365
Trp Phe Leu Arg Ala Pro Arg Arg Lys Phe Met Thr Val Ser Gly
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 146 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 99..140
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.2
seq LLLLQLSLPSPTS/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```
AAAAATGATG TCACTGGGAA CTGCAGTCAT TTGAAAAGAT AGCAATCAAG CATTTCTTTC   60

AGAGCCCTGT TCATCTTTCA GTGGCTTTGC TTCTCCTG ATG CTT TTG CTC CTT CAA  116
```

```
                                         Met Leu Leu Leu Leu Gln
                                                         -10

TTA TCT CTG CCT TCT CCC ACC TCC TCT CCG                         146
Leu Ser Leu Pro Ser Pro Thr Ser Ser Pro
            -5                       1
```

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 105 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 25..75
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.1
seq LSFKLLLLAVALG/FF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

```
AGCCCCTGCT GCTCTGGGCA GACG ATG CTG AAG ATG CTC TCC TTT AAG CTG      51
                           Met Leu Lys Met Leu Ser Phe Lys Leu
                                   -15                 -10

CTG CTG CTG GCC GTG GCT CTG GGC TTC TTT GAA GGA GAT GCT AAG TTT     99
Leu Leu Leu Ala Val Ala Leu Gly Phe Phe Glu Gly Asp Ala Lys Phe
            -5                   1                   5

GGG GAA                                                            105
Gly Glu
    10
```

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 138..203
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8
seq LLTLALLGXXXWA/GK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

```
AGCTCCTTCC TGTTCCCCTG GCGGCCCCTC GCTTCTTCCT TCTGGATGGG GGCCCAGGGG   60

GCCCAGGAGA GTATAAAGGC GATGTGGAGG GTGCCCGGCA CAACCAGACG CCCAGTCACA  120

GGGCGGAGAG CHSTGRG ATG CAC CGG CCA GAG GCC ATG CTG CTG CTG CTC     170
                   Met His Arg Pro Glu Ala Met Leu Leu Leu Leu
                           -20                 -15

ACG CTT GCC CTC CTG GGG GRC MCC AMC TGG GCA GGG AAG ATG TAT GGC    218
Thr Leu Ala Leu Leu Gly Xaa Xaa Xaa Trp Ala Gly Lys Met Tyr Gly
    -10                 -5                   1               5

CCT GGA GGA GGC AAG TAT TTC AGC ACC ACT GAA GAC TAC GAC CAT GAA    266
Pro Gly Gly Gly Lys Tyr Phe Ser Thr Thr Glu Asp Tyr Asp His Glu
                10                  15                  20

ATC ACA GGG CTG CGG GTG TCT GTA GGT CTT CTC CTG GTG AAA AGT GTC    314
Ile Thr Gly Leu Arg Val Ser Val Gly Leu Leu Leu Val Lys Ser Val
            25                  30                  35

CAG GTG AAA CTT GGA GAC TCC TGG GAC GTG                            344
Gln Val Lys Leu Gly Asp Ser Trp Asp Val
        40                  45
```

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 267 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 58..105
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8
seq VSAVLCVCAAAWC/SQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```
AAGAGCAGGG CTGCATTTCC AGCAGGAGCT GCGAGCACAG TGCTGGCTCA CAACAAG      57
```

```
ATG CTC AAG GTG TCA GCC GTA CTG TGT GTG TGT GCA GCC GCT TGG TGC     105
Met Leu Lys Val Ser Ala Val Leu Cys Val Cys Ala Ala Ala Trp Cys
    -15                 -10                     -5

AGT CAG TCT CTC GCA GCT GCC GCG GCG GTG GCT GCA GCC GGG GGG CGG     153
Ser Gln Ser Leu Ala Ala Ala Ala Ala Val Ala Ala Ala Gly Gly Arg
  1             5                  10                   15

TCG GAC GGC GGT AAT TTT CTG GAT GAT AAA CAA TGG CTC ACC ACA ATC     201
Ser Asp Gly Gly Asn Phe Leu Asp Asp Lys Gln Trp Leu Thr Thr Ile
             20                 25                  30

TCT CAG TAT GAC AAG GAA GTC GGA CAG TGG AAC AAA TTC CGA GAC GAT     249
Ser Gln Tyr Asp Lys Glu Val Gly Gln Trp Asn Lys Phe Arg Asp Asp
         35                 40                  45

GAT TAT TTC CGC ACT GGG                                             267
Asp Tyr Phe Arg Thr Gly
     50
```

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 258 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 124..174
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.8
seq VLWLISFFTFTDG/HG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```
AAGCATAAGA AGTGATTGAG CCACAAGTAT ACTGAAGGAA GGGCTCCCTC GAGTTGTGGT    60

GTGAAGAGAT AAATCACCAG TCACAGACTA TGCACCCGAC TGCTGCTGTT CAGTCCAGGG   120

AAA ATG AAA GTT GGA GTG CTG TGG CTC ATT TCT TTC TTC ACC TTC ACT     168
    Met Lys Val Gly Val Leu Trp Leu Ile Ser Phe Phe Thr Phe Thr
            -15                 -10                     -5

GAC GGC CAC GGT GGC TTC CTG GGG AAA AAT GAT GGC ATC AAA ACA AAA     216
Asp Gly His Gly Gly Phe Leu Gly Lys Asn Asp Gly Ile Lys Thr Lys
             1              5                  10

AAA GAA CTC ATT GTG AAT AAG AAA AAA CAT CTA GGC CTC GGG             258
Lys Glu Leu Ile Val Asn Lys Lys Lys His Leu Gly Leu Gly
```

```
15                    20                    25
```

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 211 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 155..202
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.7
seq ILLDLICLLFITA/CV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

```
ACTGAAATAG GAAAGTAAGA TTTATACCCA TTATTCAGCC AAAATCTGTT TTTCTTTAAC    60

TTCTACCCAT TGTTCCTAAG TCTGCCCTCT GGGGGCTGTA GAAAATAATG AAGATGATGT   120

TATTAATGAT AACCAGTGCT TGCTGTAACC AGTT ATG TGC ATT ATT TTA TTG GAT   175
                                      Met Cys Ile Ile Leu Leu Asp
                                          -15                 -10

TTA ATT TGT TTA CTC TTT ATA ACA GCA TGT GTG GGG                       211
Leu Ile Cys Leu Leu Phe Ile Thr Ala Cys Val Gly
                -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 316 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 131..307
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.6

seq FMVFGSFFPLISC/QP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

```
ACATGGATTG ATTTGTTATT TGGGGATTAA ATTAGGCAGG GCACATAGTA GGGCCTCCTT    60

GGATGTTTGA TGGCTGTTGA ATGAACGTAA GTGAATCTGT TCAGTTTTAG GGTTTTATTG   120

CATTTTTGAT ATG GAT TGT GCC AGT ATA TCT GTA AAG TTC ACT TCT ATG     169
           Met Asp Cys Ala Ser Ile Ser Val Lys Phe Thr Ser Met
                           -55                 -50

GCT ACC ATG CAT GAC TTG AGT CAG TTC TGG GCT TCT AGA GGA GAG GTT    217
Ala Thr Met His Asp Leu Ser Gln Phe Trp Ala Ser Arg Gly Glu Val
    -45                 -40                 -35

ACA AAC TGG TGG CCA GTA GGA CAA ACT AGC CTA CCA CTG TTT TAT TTG    265
Thr Asn Trp Trp Pro Val Gly Gln Thr Ser Leu Pro Leu Phe Tyr Leu
-30                 -25                 -20                 -15

GCT TTC ATG GTG TTT GGT TCT TTT TTT CCT TTA ATT TCC TGC CAG CCC    313
Ala Phe Met Val Phe Gly Ser Phe Phe Pro Leu Ile Ser Cys Gln Pro
                -10                 -5                   1

GGG                                                                316
Gly
```

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 147..206
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.6
seq LVVLFGITAGATG/AK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

```
ACTTTTGCAC TAGCAGTAGC AAGGAAGGGG GGTGGGCGCT CTTTCTTTTT CTCTTAGAAG    60

AGGGTTTAGC ACAGGTTTTT TCGTTCTCAC TTCCACACCA CCTTACCGCC TCCCGACCCC   120

CCCTCTCCCC CTCCCCACCT ATCGTC ATG ACG GCC TCT CCG GAT TAC TTG GTG   173
                             Met Thr Ala Ser Pro Asp Tyr Leu Val
                             -20                 -15
```

```
GTG CTT TTT GGG ATC ACT GCT GGG GCC ACC GGG GCC AAG CTA GGC TCG     221
Val Leu Phe Gly Ile Thr Ala Gly Ala Thr Gly Ala Lys Leu Gly Ser
    -10                 -5                   1               5

GAT GAG AAG GAG TTG ATC CTG CTG TTC TGG AAA GTC GTG GAT CTN GCC     269
Asp Glu Lys Glu Leu Ile Leu Leu Phe Trp Lys Val Val Asp Leu Ala
                10                  15                  20

AAC AAG AAG GTG GGA CAG TTG CAC GAA BKA HGN TTA GAC CGG ATC TGG     317
Asn Lys Lys Val Gly Gln Leu His Glu Xaa Xaa Leu Asp Arg Ile Trp
            25                  30                  35
```

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 46..90
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.6
seq CVLVLAAAAGAVA/VF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

```
AAGCGGCTGG TCCCCGGAAG TTGGACGCAT GCGCCGTTTC TCTGC ATG GTG TGC GTT    57
                                                  Met Val Cys Val
                                                  -15

CTC GTT CTA GCT GCG GCC GCA GGA GCT GTG GCG GTT TTC CTA ATC CTG     105
Leu Val Leu Ala Ala Ala Ala Gly Ala Val Ala Val Phe Leu Ile Leu
    -10                 -5                   1               5

CGA ATA TGG GTA GTG CTT CGT TCC ATG GAC GTT ACG CCC CGG GAG TCT     153
Arg Ile Trp Val Val Leu Arg Ser Met Asp Val Thr Pro Arg Glu Ser
                10                  15                  20

CTC AGT ATC TTG GTA GTG GCT GGG TCC GGT GGG CAT ACC ACT GAG ATC     201
Leu Ser Ile Leu Val Val Ala Gly Ser Gly Gly His Thr Thr Glu Ile
            25                  30                  35

CTG AGG CTG CTT GGG AGC TTG TCC AAT GCC TAC TCA CCT AGA CAT TAT     249
Leu Arg Leu Leu Gly Ser Leu Ser Asn Ala Tyr Ser Pro Arg His Tyr
        40                  45                  50

GTC ATT GCT GAC ACT GAT GAA ATG AGT GCC ACG                         282
Val Ile Ala Asp Thr Asp Glu Met Ser Ala Thr
```

```
55                  60
```

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 48..179
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.5
seq LMIPLLLLTPITA/TS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

```
ACAACTCAAG CCAGACAGGC AGCAATTCCA GAGTCGAAAG AGGCCTT ATG AAG AAA      56
                                                    Met Lys Lys

ACC GGG GAC GGG GGT ACT TTG AGC ACC GAG AGG ATA GGA GGG GCC GCT     104
Thr Gly Asp Gly Gly Thr Leu Ser Thr Glu Arg Ile Gly Gly Ala Ala
    -40                 -35                 -30

CTC CTC AGC CTC CTG CTG AAG AGG ATG AAG ATG ACT TTG ATG ATA CCC     152
Leu Leu Ser Leu Leu Leu Lys Arg Met Lys Met Thr Leu Met Ile Pro
-25                 -20                 -15                 -10

TTG TTG CTA TTG ACA CCT ATA ACT GCG ACC TCC ACT TCA AGG TGG CCC     200
Leu Leu Leu Leu Thr Pro Ile Thr Ala Thr Ser Thr Ser Arg Trp Pro
                 -5                   1               5

GAG ATC GGA GTA GTG GCT ATC CGC TCA CAA TTG AGG GCT TTG CAT ACC     248
Glu Ile Gly Val Val Ala Ile Arg Ser Gln Leu Arg Ala Leu His Thr
         10                  15                  20

TGT GGT CAG GAG CCC GTG CCA GCT ATG GGG TCA GAA GGG GCC GCG         293
Cys Gly Gln Glu Pro Val Pro Ala Met Gly Ser Glu Gly Ala Ala
     25                  30                  35
```

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 32..100
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.5
seq LTFLQLLLISSLP/RE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

```
AGTAGACGCT CGGGCACCAG CMGCGGCAAG G ATG GAG CTG GGT TGC TGG ACG        52
                                   Met Glu Leu Gly Cys Trp Thr
                                                   -20

CAG TTG GGG CTC ACT TTT CTT CAG CTC CTT CTC ATC TCG TCC TTG CCA     100
Gln Leu Gly Leu Thr Phe Leu Gln Leu Leu Leu Ile Ser Ser Leu Pro
    -15                 -10                     -5

AGA GAG TAC ACA GTC ATT AAT GAA GCC TGC CCT GGA GCA GAG TGG AMT     148
Arg Glu Tyr Thr Val Ile Asn Glu Ala Cys Pro Gly Ala Glu Trp Xaa
  1               5                  10                  15

ATC ATG TGT CGG GAG TGC TGT GAA TAT GAT CAG ATT GAG TGC GTC TGC     196
Ile Met Cys Arg Glu Cys Cys Glu Tyr Asp Gln Ile Glu Cys Val Cys
              20                  25                  30

CCC GGA AAG AGG GAA GTC GTG GGT TAT ACC ATC CCT TGC TGC AGG AAT     244
Pro Gly Lys Arg Glu Val Val Gly Tyr Thr Ile Pro Cys Cys Arg Asn
          35                  40                  45

GAG GMG AAT GAG TGT GAC TCC TGC CTG ATC CAC CCA GGT TGT ACC ATC     292
Glu Xaa Asn Glu Cys Asp Ser Cys Leu Ile His Pro Gly Cys Thr Ile
      50                  55                  60

TTT GAA AAC TGC AMG AGC TGC CGM AAT GGC TCA TGG GGG GGT ACC TTG     340
Phe Glu Asn Cys Xaa Ser Cys Arg Asn Gly Ser Trp Gly Gly Thr Leu
 65                  70                  75                  80
```

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 351 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:

(A) NAME/KEY: sig_peptide
(B) LOCATION: 112..192
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.2
seq SLLFFLLLEGGXT/EQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

```
AAGACCTCGG AACGAGAGCG CCCCGGGGAG CTCGGAGCGC GTGCACGCGT GGCAVACGGA    60

GAAGGCVAKK RCNNNNRCTT GAAGGTTCTG TCACCTTTTG CAGTGGTCCA A ATG AGA   117
                                                         Met Arg

RAA AAG TGG AAA ATG GGA GGC ATG AAA TAC ATC TTT TCG TTG TTG TTC      165
Xaa Lys Trp Lys Met Gly Gly Met Lys Tyr Ile Phe Ser Leu Leu Phe
-25                 -20                 -15                 -10

TTT CTT TTG CTA GAA GGA GGC KAA ACA GAG CAA GTR AMN CAT TCA GAG      213
Phe Leu Leu Leu Glu Gly Gly Xaa Thr Glu Gln Val Xaa His Ser Glu
                 -5                   1               5

ACA TAT TGC ATG TTT CAA GAC AAG AAG TAC AGA GTG GGT GAG AGA TGG      261
Thr Tyr Cys Met Phe Gln Asp Lys Lys Tyr Arg Val Gly Glu Arg Trp
        10                  15                  20

CAT CCT TAC CTG GAA CCT TAT GGG TTG GTT TAC TGC GTG AAC TGC ATC      309
His Pro Tyr Leu Glu Pro Tyr Gly Leu Val Tyr Cys Val Asn Cys Ile
    25                  30                  35

TGC TCA GAG RAT GGG AAT GTG CTT TGC AGC CGA GTC AGA TGT              351
Cys Ser Glu Xaa Gly Asn Val Leu Cys Ser Arg Val Arg Cys
 40                  45                  50
```

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 68..124
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.2
seq VSIMLLLVTVSDC/AV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

```
AAAGGCTGAG CGGGAGGAAG CGAGAGGCAT CTAAGCAGGC AGTGTTTTGC CTTCACCCCA    60
```

```
AGTGACC ATG AGA GGT GCC ACG CGA GTC TCA ATC ATG CTC CTC CTA GTA     109
        Met Arg Gly Ala Thr Arg Val Ser Ile Met Leu Leu Leu Val
                        -15                 -10

ACT GTG TCT GAC TGT GCT GTG ATC ACA GGG GCC TGT GAG CGG GAT GTC     157
Thr Val Ser Asp Cys Ala Val Ile Thr Gly Ala Cys Glu Arg Asp Val
 -5                 1               5                   10

CAG TGT GGG GCA GGC ACC TGC TGT GCC ATC AGC CTG TGG CTT CGA GGG     205
Gln Cys Gly Ala Gly Thr Cys Cys Ala Ile Ser Leu Trp Leu Arg Gly
            15                  20                  25

CTG CGG ATG TGC ACC CCG CTG GGG CGG GAA GGC GAG GAG TGC CAC CCC     253
Leu Arg Met Cys Thr Pro Leu Gly Arg Glu Gly Glu Glu Cys His Pro
        30                  35                  40

GGC AGC CAC AAG ATC CCC TTC TTC AGG AAA CGC AAG CAC CAC ACC TGT     301
Gly Ser His Lys Ile Pro Phe Phe Arg Lys Arg Lys His His Thr Cys
    45                  50                  55

CCT TGC TTG                                                         310
Pro Cys Leu
 60
```

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 380 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 240..302
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.2
seq SALLFSLLCEAST/VV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

```
ACCTTTCTGG ACGTTGCAAA CTGTGACATA TAAAAGCTGT TAGCTGCTCC TCTAGCCAGC    60

AGCATTCAAA CCTTGCAGAG CTTTGCTCTC AGAGAGTTTG TAAAAAGACA CACTCCTCTT   120

ACAAGAGTTC ATGCTACCAC ATAGCAAAGA ACCTTAAATT TTTGGAAGAA CAATATATTC   180

ATTTTGGCAT TGTGCAGAGC AAAGTAAACT CGGTGGCCTC TTCTTCTCCA CCCCTCAAR    239

ATG ATA GCR ATC TCT GCC GTC AGC AGT GCA CTC CTG TTC TCC CTT CTC     287
Met Ile Ala Ile Ser Ala Val Ser Ser Ala Leu Leu Phe Ser Leu Leu
```

```
-20                    -15                    -10

TGT GAA GCA AGT ACC GTC GTC CTA CTC AAT TCC ACT GAC TCA TCC CCG     335
Cys Glu Ala Ser Thr Val Val Leu Leu Asn Ser Thr Asp Ser Ser Pro
 -5                  1                   5                  10

CSA ACC AAT AAT TTC RCT GAT AWT GAA GCA GCT CTG AAA GCA CAT         380
Xaa Thr Asn Asn Phe Xaa Asp Xaa Glu Ala Ala Leu Lys Ala His
             15                  20                  25
```

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 435 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 181..243
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.2
seq SALLFSLLCEAST/VV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

```
AGCATTCAAA CCTTGCAGAG CTTTGCTCTC AGAGAGTTTG TAAAAAGACA CACTCCTCTT    60

ACAAGAGTTC ATGCTACCAC ATAGCAAAGA ACCTTAAATT TTTGGAAGAA CAATATATTC   120

MATTTTGGCA TTGTGCAGAG CAAAGTAAAC TCGGTGGCCT CTTCTTCTCC ACCCCTCAAA   180

ATG ATA GCA ATC TCT GCC GTC AGC AGT GCA CTC CTG TTC TCC CTT CTC     228
Met Ile Ala Ile Ser Ala Val Ser Ser Ala Leu Leu Phe Ser Leu Leu
    -20                 -15                 -10

TGT GAA GCA AGT ACC GTC GTC CTA CTC AAT TCC ACT GAC TCA TCC CCG     276
Cys Glu Ala Ser Thr Val Val Leu Leu Asn Ser Thr Asp Ser Ser Pro
 -5                  1                   5                  10

CCA ACC AAT AAT TTC ACT GAT ATT GAA GCA GCT CTG AAA GCA CAA TTA     324
Pro Thr Asn Asn Phe Thr Asp Ile Glu Ala Ala Leu Lys Ala Gln Leu
             15                  20                  25

GAT TCA GCG GAT ATC CCC AAA GCC AGG CGG AAG CGC TAC ATT TCG CAG     372
Asp Ser Ala Asp Ile Pro Lys Ala Arg Arg Lys Arg Tyr Ile Ser Gln
         30                  35                  40

AAT GAC ATG ATC GCC ATT CTT GAT TAT CAT AAT CAA GTT CGG GGC AAA     420
Asn Asp Met Ile Ala Ile Leu Asp Tyr His Asn Gln Val Arg Gly Lys
     45                  50                  55
```

```
GTG TTC CCA MCG GCA                                                   435
Val Phe Pro Xaa Ala
 60
```

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 426 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 352..417
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.2
seq LLTLVLCVAVAYE/RQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

```
ATTGAGCTGT CTGCAGCAGA GCTGAGAGGA CCAGCCATTT TACTTATGGA AAACAGTGTG   60

GCATATTCTG CTGAGCTTCG CCCTGGAAGA AGCCTCTTTT ATACATCTCT TCAGGGAAGA  120

GAGAAGCAAT GGGCATGTTA GTATACAATG ATCACAGCCA CGCAGGCCTG CAAGCTGCCT  180

TTTGGACAGG CTGTTGACTG CCGTTCCAAT TAGCTGATTG GAGAATGTGG AATGCAGAGT  240

GATAATGCTG CATATCTGCT ATCAGGCAGC AGCAAAGGTT TTTGTCTTGG GAAGGCAAGC  300

TTTCCCTGCA ATATTATCTC AGCAGCTCCC TAGCTGCTTA CCCTGAAAAC G ATG GAT   357
                                                          Met Asp

CCA AAC GGA GGG TGT TGC ACT CTG CTA ACG CTG GTC CTG TGC GTG GCT    405
Pro Asn Gly Gly Cys Cys Thr Leu Leu Thr Leu Val Leu Cys Val Ala
-20                 -15                 -10                  -5

GTG GCA TAT GAG CGG CAG GAG                                       426
Val Ala Tyr Glu Arg Gln Glu
                 1
```

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 389 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 288..362
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.2
seq LFTFSTSLPSSLS/SS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

```
ACAATACCTG TTACTTATAT ACTTTTCTTT GTCTAAAAAA GAAATAAGAT CTGTCTAGAT    60

GACTGATTAA CTTAGGGAGA TTCTGATTAA CAGAATTTCT AGAAATGGCT TTCAGCAGGC   120

AAAGAGAAAA TTATATTTTG TACCAATTTA TATAAAGTTC ATCTAGCTCA GCTTTTGGAG   180

ATGTCCCTGG GGCTAGAGAT GAAATATCGT TTTCCTGTCC ACAGACAGCG GTCTGCAGTT   240

CACCCCATGA ACTCATACAG GTCAGAATTA AACCCCGAGC TTTGTTT ATG GAG GGT     296
                                                    Met Glu Gly
                                                    -25

GAG ATA TAT TTC CAA GTA TTT CTT TCT CTT TTC ACA TTT TCC ACA TCA     344
Glu Ile Tyr Phe Gln Val Phe Leu Ser Leu Phe Thr Phe Ser Thr Ser
        -20                 -15                 -10

TTA CCA TCA TCA TTG TCG TCA TCA TCA TTG TCA TCA TCC AAT GGG         389
Leu Pro Ser Ser Leu Ser Ser Ser Ser Leu Ser Ser Ser Asn Gly
    -5                  1               5
```

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 328 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 194..316
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7
seq FLCMLAAIDLALS/TS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

```
ATGAGTCAGC CTGAAAGGAA CAGGCCGAAC TGCTGTATGG GCTCTACTGC CAGTGTGACC   60

TCACCCTCTC CAGTCACCCC TCCTCAGTTC CAGCTATGAG TTCCTGCAAC TTCACACATG  120

CCACCTTTGT GCTTAATKGG AATCCCAGGG ATTAGAGAAA GCCCATTTCT GGGTTGGCTT  180

CCCCCTCCTT TCC ATG TAT GTA GTG GCA ATG TTT GGA AAC TGC ATC GTG     229
               Met Tyr Val Val Ala Met Phe Gly Asn Cys Ile Val
                   -40                 -35                 -30

GTC TTC ATC GTA AGG ACG GAA CGC AGC CTG CAC GCT CCG ATG TAC CTC    277
Val Phe Ile Val Arg Thr Glu Arg Ser Leu His Ala Pro Met Tyr Leu
                -25                 -20                 -15

TTT CTC TGC ATG CTT GCA GCC ATT GAC CTG GCC TTA TCC ACA TCC ACC    325
Phe Leu Cys Met Leu Ala Ala Ile Asp Leu Ala Leu Ser Thr Ser Thr
            -10                 -5                  1

ATG                                                                328
Met
```

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 267 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 79..207
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7
seq PWFLAPWCPGTQS/NR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

```
ACCCTTCGTT CTGGTTCTGG TTCTAGTTCT GGTTCTAACA ACTCACAATC CCTTTAGCTT   60

TCTCTCCCCT CCCTTTGA ATG AGA GAA ACT AMC CCG CTT CCG AAG CCC CTG    111
                    Met Arg Glu Thr Xaa Pro Leu Pro Lys Pro Leu
                                -40                 -35

AAA GAC ACT GCT CCT TCC TCT CAT GGA GTT GGC TCC GAC AGC CCG TCT    159
Lys Asp Thr Ala Pro Ser Ser His Gly Val Gly Ser Asp Ser Pro Ser
        -30                 -25                 -20

GCC ACC AGG CCA TGG TTC CTT GCC CCA TGG TGT CCT GGG ACC CAG AGC    207
Ala Thr Arg Pro Trp Phe Leu Ala Pro Trp Cys Pro Gly Thr Gln Ser
```

```
-15                 -10                -5

AAC AGG ATC TGT CAC CCA CCT CTC TCT TCT CCC CCA GAT CAA GCG ACG      255
Asn Arg Ile Cys His Pro Pro Leu Ser Ser Pro Pro Asp Gln Ala Thr
  1               5                  10                  15

TGC CTC AGA GGC                                                      267
Cys Leu Arg Gly
             20
```

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 301 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 23..202
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7
seq VLVVLALRSLGRS/CS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

```
AAGTGAGGCT TGGAAAGGCG TC ATG GAC AGA CCT GGG TCG CTT TCT GTC TTC      52
                         Met Asp Arg Pro Gly Ser Leu Ser Val Phe
                         -60                 -55

GGG TCC CTC CCG GCT TCG CTC GGG ACC TGG CTC TCA AGC CCA GCT TGG      100
Gly Ser Leu Pro Ala Ser Leu Gly Thr Trp Leu Ser Ser Pro Ala Trp
-50                 -45                 -40                 -35

CTG GTG GAC AGA CCG GTG CGC TCT GCA CAC CCG AGT GCG AAT TCC ACC      148
Leu Val Asp Arg Pro Val Arg Ser Ala His Pro Ser Ala Asn Ser Thr
                -30                 -25                 -20

GGC GTG AGA ATG AGC GTG CTC GTG GTC CTG GCC CTG AGG TCC CTG GGT      196
Gly Val Arg Met Ser Val Leu Val Val Leu Ala Leu Arg Ser Leu Gly
            -15                 -10                 -5

CGC AGC TGT TCC CTC TCC CAG GCT GCC CCC TCC AGG TGG ACG CGG TCA      244
Arg Ser Cys Ser Leu Ser Gln Ala Ala Pro Ser Arg Trp Thr Arg Ser
         1                   5                  10

AAC GAT GCC CCG CAG CCT CCT GGG TCT CAG CAC ATA TTC CAC ACC TAH      292
Asn Asp Ala Pro Gln Pro Pro Gly Ser Gln His Ile Phe His Thr Xaa
 15                  20                  25                  30

GTS CCC GGG                                                          301
```

Val Pro Gly

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 110 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 3..65
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7
seq VILLFSYPSCCLC/FL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

```
AT ATG CAT TAT TTT GTT GCT GGG AAA GTA ATC CTT CTC TTC TCT TAT       47
   Met His Tyr Phe Val Ala Gly Lys Val Ile Leu Leu Phe Ser Tyr
       -20                 -15                 -10

CCA TCA TGT TGT TTG TGT TTC TTG GTG TAC AGG AGA GTA AGC WAT TTA      95
Pro Ser Cys Cys Leu Cys Phe Leu Val Tyr Arg Arg Val Ser Xaa Leu
     -5                  1               5                  10

TTT AAG TGC TTT GAG                                                 110
Phe Lys Cys Phe Glu
                 15
```

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 318 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 160..216
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7

seq STVVLQVLTQATS/QD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

```
AGACGCCARA CATGGCGTGT TCCTAGAAGC CGCTTTCGGC ATCAGTAGGC GGCGGCGTGG    60

GGTCTGGCAK CGTGGGGAGA GGGAMCAACC GACGCCACTT CGTGTTGGGA AGTGGGAGCG   120

GGANRGCCGG GCAATTCCCG ACCGAACCAA ACGGTTTCC ATG GAT CTC AAT AGT      174
                                           Met Asp Leu Asn Ser
                                                           -15

GCC AGC ACT GTT GTT CTT CAG GTG TTA ACA CAG GCC ACC AGT CAG GAT     222
Ala Ser Thr Val Val Leu Gln Val Leu Thr Gln Ala Thr Ser Gln Asp
                -10                 -5                   1

ACT GCT GTG TTA AAA CCA GCT GAG GAG CAG TTG AAG CAG TGG GAG ACA     270
Thr Ala Val Leu Lys Pro Ala Glu Glu Gln Leu Lys Gln Trp Glu Thr
         5                  10                  15

CAG CCA GGW TTC TAT TCA GTG TTG CTG AAT ATT TTC ACC AAC CAC GGG     318
Gln Pro Gly Phe Tyr Ser Val Leu Leu Asn Ile Phe Thr Asn His Gly
    20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 95..313
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7
seq FLCMLAAIDLALS/TS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

```
ATGAGTCAGC CTGAAAGAAC AGGCCGAACT GCTGTATGGG CTCTACTGCC AGTGTGACCT    60

CACCCTCTCC AGTCACCCCT CCTCAGTTCC AGCT ATG AGT TCC TGC AAC TTC ACA  115
                                      Met Ser Ser Cys Asn Phe Thr
                                                      -70

CAT GCC ACC TTT GTG CTT ATT GGT ATC CCA GGA TTA GAG AAA GCC CAT     163
His Ala Thr Phe Val Leu Ile Gly Ile Pro Gly Leu Glu Lys Ala His
    -65                 -60                 -55
```

```
TTC TGG GTT GGC TTC CCC CTC CTT TCC ATG TAT GTA GTG GCA ATG TTT     211
Phe Trp Val Gly Phe Pro Leu Leu Ser Met Tyr Val Val Ala Met Phe
-50                 -45                 -40                 -35

GGA AAC TGC ATC GTG GTC TTC ATC GTA AGG ACG GAA CGC AGC CTG CAC     259
Gly Asn Cys Ile Val Val Phe Ile Val Arg Thr Glu Arg Ser Leu His
                -30                 -25                 -20

GCT CCG ATG TAC CTC TTT CTC TGC ATG CTT GCA GCC ATT GAC CTG GCC     307
Ala Pro Met Tyr Leu Phe Leu Cys Met Leu Ala Ala Ile Asp Leu Ala
            -15                 -10                 -5

TTA TCC ACA TCC ACC ATG                                             325
Leu Ser Thr Ser Thr Met
        1
```

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 179..346
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 6.9
 seq PLFFSCSISATHS/CV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

```
ACAAAATCAA GAAAATCCAA CATAGATGGT CAAAATATTC ATAGGTGACT GAGAGTATCC  60

AAATGGGCCA GGTGACTGAG AATACGCAAA CAGGCCAGAA TAATATCTGT GTTAAATTTG  120

ACCCTCTATT TTATTAACAT ATCTGTCATG ACCTTTCTCT GTACCTGCTG TAGTACTC    178

ATG TAT AGA CTC AGT CTT ATA GCA GGC CCT GGG TCC TAT CCT GTG CTA     226
Met Tyr Arg Leu Ser Leu Ile Ala Gly Pro Gly Ser Tyr Pro Val Leu
    -55                 -50                 -45

AGA TGG GGA GTT TGG GAC ATC CCT AGT TCA TTA GTT CAA GTG ACT TAC     274
Arg Trp Gly Val Trp Asp Ile Pro Ser Ser Leu Val Gln Val Thr Tyr
-40                 -35                 -30                 -25

CAT CAG CCC AAC CTC ACT ACA AAT TTG GAT CTG CCT TTG TTC TTC AGT     322
His Gln Pro Asn Leu Thr Thr Asn Leu Asp Leu Pro Leu Phe Phe Ser
                -20                 -15                 -10

TGT AGT ATC TCG GCT ACC CAT TCT TGT GTC AAG CCT CCA TCT GTA ATT     370
```

```
Cys Ser Ile Ser Ala Thr His Ser Cys Val Lys Pro Pro Ser Val Ile
            -5                   1                   5

ATT GGT ATC TCT TCT TTC CTG AGC TTT CCT TAT CAA ACT TTG GTA         415
Ile Gly Ile Ser Ser Phe Leu Ser Phe Pro Tyr Gln Thr Leu Val
     10                  15                  20
```

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 400 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 128..199
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.9
seq LCFLLLAVAMSFF/GS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

```
AAGTTGGTGA GCTTTTCCGG TGCTCTGCAC AGATGCTGGG GCGCTGAGCA AACAGCCCTC    60

AGTTTCTGGA GCTGTTCCGA GTCCCGTGGA GTCTCCATCT GAGCCCTTTC CTAGTCCAGG   120

CATCCCG ATG TTG GTG GAT GGC CCA TCT GAG CGG CCA GCC CTG TGC TTC      169
        Met Leu Val Asp Gly Pro Ser Glu Arg Pro Ala Leu Cys Phe
                        -20                 -15

TTG CTG TTG GCT GTG GCA ATG TCT TTC TTC GGC TCA GCT CTA TCC ATA      217
Leu Leu Leu Ala Val Ala Met Ser Phe Phe Gly Ser Ala Leu Ser Ile
-10                 -5                   1                   5

GAT GAA ACA CGG GCG CAT CTG TTG TTG AAA GAA AAG ATG ATG CGG CTG      265
Asp Glu Thr Arg Ala His Leu Leu Leu Lys Glu Lys Met Met Arg Leu
            10                  15                  20

GGG GGG CGG CTG GTG CTG AAC ACC AAG GAG GAG CTG GCC AAT GAG AGG      313
Gly Gly Arg Leu Val Leu Asn Thr Lys Glu Glu Leu Ala Asn Glu Arg
        25                  30                  35

CTC ATG ACG CTC AAA ATC GCT GAG ATG AAG GAG GCC ATG AGG ACC CTG      361
Leu Met Thr Leu Lys Ile Ala Glu Met Lys Glu Ala Met Arg Thr Leu
    40                  45                  50

ATA TTC CCA CCC AGC ATG CAC TTT TTC CAG GCC AAA TGG                  400
Ile Phe Pro Pro Ser Met His Phe Phe Gln Ala Lys Trp
 55                  60                  65
```

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 33..137
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.9
seq XLXXLLTPPPSYG/HQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

```
AACCGGCCCG CGCCCCGCCA TGGAGGACCT GG ATG CCC TGC TCT CTG ACC TGG      53
                                    Met Pro Cys Ser Leu Thr Trp
                                    -35                 -30

AGA CTA CCA CCT CGC ACA TGC CAA KGT CAK GGG CTY CYA AAG AGC GYY     101
Arg Leu Pro Pro Arg Thr Cys Gln Xaa Xaa Gly Leu Xaa Lys Ser Xaa
            -25                 -20                 -15

CTT GYG GAB CTK CTC ACC CCT CCC CCA TCC TAT GGC CAC CAG CCA CAG     149
Leu Xaa Xaa Leu Leu Thr Pro Pro Pro Ser Tyr Gly His Gln Pro Gln
        -10                 -5                  1

ACA GGG TCT GGG GAG TCT DCA GGA GCC TCG GGG GAC AAG GAC CAC CTG     197
Thr Gly Ser Gly Glu Ser Xaa Gly Ala Ser Gly Asp Lys Asp His Leu
  5                 10                  15                  20

TAC AGC ACG GTA TGC                                                 212
Tyr Ser Thr Val Cys
                 25
```

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 15..137
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.8
seq LFLFLTSIAEXCS/TP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

```
ACCCTGTKCT TKTC ATG GTT DTC TGG CTC GTC TTA TTT GCT CTT CAG ATT      50
                Met Val Xaa Trp Leu Val Leu Phe Ala Leu Gln Ile
                    -40                 -35                 -30

TAC TCC TAT KKY AGT ACT CGA GAT CAG CCT GCA TCA CGT GAK AGG CTT      98
Tyr Ser Tyr Xaa Ser Thr Arg Asp Gln Pro Ala Ser Arg Xaa Arg Leu
                -25                 -20                 -15

CTT TTC CTT TTT CTG ACA AGT ATT GCG GAA TRC TGC AGC ACT CCT TAC     146
Leu Phe Leu Phe Leu Thr Ser Ile Ala Glu Xaa Cys Ser Thr Pro Tyr
            -10                  -5                   1

TCT CTT TTG GGT TTK GTC TTC ACG GTT TCT TTT GTT GCC TTG GGT GTT     194
Ser Leu Leu Gly Xaa Val Phe Thr Val Ser Phe Val Ala Leu Gly Val
      5                  10                  15

CTC ACA CTC TGC AAG TTT TAC TTG CAG GGT TAT CGA GCT TTC ATG AAT     242
Leu Thr Leu Cys Lys Phe Tyr Leu Gln Gly Tyr Arg Ala Phe Met Asn
 20                  25                  30                  35

GAT CCT GCC ATG AAT CGG GGA GGT GCG                                 269
Asp Pro Ala Met Asn Arg Gly Gly Ala
                40
```

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 9..62
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.7
seq LPLLXXXSLPVGA/WL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

```
AAGTCCTG ATG GCC CGG CAT GGG TTA CCG CTG CTG CHB YWG HTG TCG CTC     50
         Met Ala Arg His Gly Leu Pro Leu Leu Xaa Xaa Xaa Ser Leu
                     -15                 -10                 -5

CCG GTC GGC GCG TGG CTC                                              68
Pro Val Gly Ala Trp Leu
                      1
```

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 407 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 258..368
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.7
seq ILYILWYCSVCSS/GS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

```
AAGGTTGGTC TGGACCGGAA GCGAAGATGG CGACTTCTGG CGCGGCCTCG GCGGASTGGT    60

GATCGGCTGG TGCATATTCG GCCTCTTACT ACTGGCKATT TTGGCATTCT GCTGGATATA   120

TGTTCGTAAA TACCAAAGTC GGCGGGAAAG TGAAGTTGTC TCCACCATAA CAGCAATTTT   180

TTCTCTAGCA ATTGCACTTA TCACATCAGC ACTTCTACCA GTGGATATAT TTTTGGTTTC   240

TTACATGAAA AATCAAA ATG GTA CAT TTA AGG ACT GGG CTA ATG CTA ATG      290
                   Met Val His Leu Arg Thr Gly Leu Met Leu Met
                           -35                 -30

TCA GCA GAC AGA TTG AGG ACA CTG TAT TAT ACG GTT ACT ATA CTT TAT      338
Ser Ala Asp Arg Leu Arg Thr Leu Tyr Tyr Thr Val Thr Ile Leu Tyr
    -25                 -20                 -15

ATT CTG TGG TAT TGT TCT GTG TGT TCT TCT GGA TCC CTT TTG TCT ACT      386
Ile Leu Trp Tyr Cys Ser Val Cys Ser Ser Gly Ser Leu Leu Ser Thr
-10                 -5                   1                   5

TCT ATT ATG AAG AAA AGG ATG                                          407
Ser Ile Met Lys Lys Arg Met
            10
```

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 348 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 196..240
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.7
seq ILSTVTALTFARA/LD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

```
AAAAAATTGG TCCCAGTTTT CACCCTGCCG CAGGGCTGGC TGGGGAGGGC AGCGGTTTAG    60

ATTAGCCGTG GCCTAGGCCG TTTAACGGGG TGACACGAGC HTGCAGGGCC GAGTCCAAGG   120

CCCGGAGATA GGACCAACCG TCAGGAATGC GAGGAATGTT TTTCTTCGGA CTCTATCGAG   180

GCACACAGAC AGACC ATG GGG ATT CTG TCT ACA GTG ACA GCB TTA ACA TTT   231
                 Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
                 -15                 -10                 -5

GCC AGA GCC CTG GAC GGC TGC AGA AAT GGC ATT GCC CAC CCT GCA AGT    279
Ala Arg Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
            1               5                  10

GAG AAG CAC AGA CTC GAG AAA TGT AGG GAA CTC GAG AGC AGC CAC TCG    327
Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Ser Ser His Ser
    15                  20                  25

GCC CCA GGA TCA ACC CAG CAG
Ala Pro Gly Ser Thr Gln Gln                                        348
 30                 35
```

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 146 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

```
      (ix) FEATURE:
            (A) NAME/KEY: sig_peptide
            (B) LOCATION: 45..113
            (C) IDENTIFICATION METHOD: Von Heijne matrix
            (D) OTHER INFORMATION:  score 6.5
                                    seq LTFLQXLLISSLX/RE

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

ACTCTCCCTC CCCAGTAGAC GCTCGGGCAC CAGCCGCGGC AAGG ATG GAG CTG GGT    56
                                                 Met Glu Leu Gly
                                                             -20

TGC TGG ACG CAG TTG GGG CTC ACT TTT CTT CAG STC CTT CTC ATC TCG    104
Cys Trp Thr Gln Leu Gly Leu Thr Phe Leu Gln Xaa Leu Leu Ile Ser
                -15                 -10                 -5

TCC TTG CHA AGA GAG TAC ACA GTC ATT AAT GAA GCH CGC AAG            146
Ser Leu Xaa Arg Glu Tyr Thr Val Ile Asn Glu Ala Arg Lys
             1               5                  10


(2) INFORMATION FOR SEQ ID NO: 86:

      (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 308 base pairs
            (B) TYPE: NUCLEIC ACID
            (C) STRANDEDNESS: DOUBLE
            (D) TOPOLOGY: LINEAR

      (ii) MOLECULE TYPE: CDNA

      (vi) ORIGINAL SOURCE:
            (A) ORGANISM: Homo Sapiens
            (F) TISSUE TYPE: Prostate

      (ix) FEATURE:
            (A) NAME/KEY: sig_peptide
            (B) LOCATION: 201..266
            (C) IDENTIFICATION METHOD: Von Heijne matrix
            (D) OTHER INFORMATION:  score 6.4
                                    seq FLLCXSVFTDCKG/DV

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

ACAGAATCAC GTTTTDAGTT GTGCGTGTGC GCGCACACGM GTGTAAAMAG CACTTTCGAT   60

TGTGCCTCCT GTTTTCTCGA GTGGGGACAC TTTAACTACA GTTTASACCT CGGGCGCATM  120

AAGTTTKTCT TCTCTTTCTC TCTGGTTRTT TCTGTTTCTG AGTGGACCAA CAGCAGARCC  180

CACGAGGAKT TGTTTTGAGT ATG GAG CTG TTG CGG GTD TGC TCC TTT TTC TTG  233
                      Met Glu Leu Leu Arg Val Cys Ser Phe Phe Leu
                              -20                 -15

CTT TGC TSC TCA GTT TTT ACA GAC TGT AAA GGA GAT GTG TTG TGT GTG    281
Leu Cys Xaa Ser Val Phe Thr Asp Cys Lys Gly Asp Val Leu Cys Val
    -10                 -5                   1               5
```

```
AAG ATG GAG CAG AGT CAA ATC TGT GCT                                  308
Lys Met Glu Gln Ser Gln Ile Cys Ala
                10
```

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 289 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 203..268
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.3
seq TWFLLLPPGQCRA/VG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

```
AGAATCTCAC GAGAGAAGAA AACCAGCCAC ATAAAGGATT TGAAAGCTCA ACTTGCTTTC    60

CCACTCTGTT ATCCCTGGAG TTGGCTTGGA TTCACCCTGA AGCCTTCCCC CTCCCGGGGA   120

AAGTTGCTTC ACGTTGCAGC TCAGCAGGTT TGTCCAGCTA CATAGGCTCC AGAAAACAAG   180

AAGCAAGACT GGAAAGCTGG GG ATG ATT GTA CGC CCT CGC CTG AAT CTT ACG     232
                         Met Ile Val Arg Pro Arg Leu Asn Leu Thr
                                 -20                     -15

TGG TTC CTC CTT CTT CCA CCT GGC CAG TGC AGA GCC GTG GGT GCC ACG      280
Trp Phe Leu Leu Leu Pro Pro Gly Gln Cys Arg Ala Val Gly Ala Thr
        -10                 -5                       1

TGG CCC GGG                                                          289
Trp Pro Gly
  5
```

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 120 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..57
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.3
seq MVALCCCLWKISG/CE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

```
ATG CAA TTC TTG TTT AAG ATG GTG GCC TTA TGC TGT TGT CTC TGG AAG      48
Met Gln Phe Leu Phe Lys Met Val Ala Leu Cys Cys Cys Leu Trp Lys
                -15                 -10                 -5

ATC TCC GGC TGT GAG GAA GTC CCT CTA ACT TAC AAC CTG CTC AAG TGC      96
Ile Ser Gly Cys Glu Glu Val Pro Leu Thr Tyr Asn Leu Leu Lys Cys
              1               5                  10

CTC CTA GAT AAA GCG CAC GTA GGG                                     120
Leu Leu Asp Lys Ala His Val Gly
     15                  20
```

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 247 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 50..112
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.3
seq CVCAAAXXSQSLX/XX

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

```
AAAGCGTCCT ATCCGGAGCC AACTGTAGCT GGGATCCAGC GAGAGGAAG ATG CTC AAG     58
                                                      Met Leu Lys
                                                          -20

GTG TCA GCC GTA CTG TGT GTG TGT GCA GCC GCT TDG TGS AGT CAG TCT     106
Val Ser Ala Val Leu Cys Val Cys Ala Ala Ala Xaa Xaa Ser Gln Ser
            -15                 -10                     -5

CTC GSM RCT KCC GCG GCG GTG GCT GCA GCC GGG GGG CGG TCG GAC GGC     154
```

```
Leu Xaa Xaa Xaa Ala Ala Val Ala Ala Ala Gly Gly Arg Ser Asp Gly
          1                 5                       10

GGT AAT TTT CTG GAT GAT AAA CAA TGG CTC ACC ASR ATC TCT CAG TAT     202
Gly Asn Phe Leu Asp Asp Lys Gln Trp Leu Thr Xaa Ile Ser Gln Tyr
 15                 20                   25                  30

GAC AAG GAA KTC GGM MAG TGG AAC AAA TTC CGA GAC GAT KAT TAT         247
Asp Lys Glu Xaa Gly Xaa Trp Asn Lys Phe Arg Asp Asp Xaa Tyr
                35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 294 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 124..186
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.3
seq MVALCCCLWKISG/CE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

```
AAGACGCTGC CTTTAGGGAG AGATAAAAAG CATAATGACA TTAGCTAGGA AAGTTAATTT   60

TCAGTTCTTA CTGAAGTGCT GTATGAAACT GAAATTTCCA AGGAACTGAA TTTTGTGAGC  120

CAA ATG AGC ATG CAA TTC TTG TTT AAG ATG GTG GCC TTA TGC TGT TGT     168
    Met Ser Met Gln Phe Leu Phe Lys Met Val Ala Leu Cys Cys Cys
        -20                 -15                 -10

CTC TGG AAG ATC TCC GGC TGT GAG GAA GTC CCT CTA ACT TAC AAC CTG     216
Leu Trp Lys Ile Ser Gly Cys Glu Glu Val Pro Leu Thr Tyr Asn Leu
     -5                  1                   5                  10

CTC AAG TGC CTC CTA GAT AAA GCG CAC TGT GTA CTC CTG ACA CCT TGT     264
Leu Lys Cys Leu Leu Asp Lys Ala His Cys Val Leu Leu Thr Pro Cys
                15                  20                  25

GGT TAC ATC TTT TCC TTG ATC AGT CCA GGG                             294
Gly Tyr Ile Phe Ser Leu Ile Ser Pro Gly
            30                  35
```

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 114..164
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.2
seq LWILLGSLSCRTS/NR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

```
AATTCTTATA GGTGTGTCCA GCAGGCAGTG GCTTGTAGCT GTTCCTTCAG CCACTTAACA    60

GGTTTGATTT CAAAGCTTTT TAATAGAGAA ACTAACATGT TTGGAGGGGA TTC ATG      116
                                                           Met

GCC CAA CAT TTA TGG ATT TTG TTG GGA AGT CTC AGT TGC CGA ACA AGC     164
Ala Gln His Leu Trp Ile Leu Leu Gly Ser Leu Ser Cys Arg Thr Ser
    -15                 -10                     -5

AAC CGG CGG                                                         173
Asn Arg Arg
  1
```

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 66..149
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.1
seq LYLFSGFWTFXLG/KF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

```
ACACTTGART TGGGGTTAAG TTGAAGAACA GACAAACTTA GACACAAAGC TATGCAAAAA   60

TTGTG ATG AAC AAG GAA RAA GTA AGT TTN GAA AGG ARA GCA CAG GTC AGA  110
      Met Asn Lys Glu Xaa Val Ser Xaa Glu Arg Xaa Ala Gln Val Arg
                  -25                 -20                 -15

TTA TAT TTA TTC TCA GGA TTT TGG ACT TTT KTA TTA GGG AAA TTT AAA    158
Leu Tyr Leu Phe Ser Gly Phe Trp Thr Phe Xaa Leu Gly Lys Phe Lys
            -10                  -5                   1

CAA GGG GAA TGR TCT TAT ATK KGT ATT CTA GAA AGA TTA CTG TGG CAG    206
Gln Gly Glu Xaa Ser Tyr Xaa Xaa Ile Leu Glu Arg Leu Leu Trp Gln
        5                  10                  15

CAG CAG TAT GWA GGA TGG CTT GTA GGR GAT AAG AGA                    242
Gln Gln Tyr Xaa Gly Trp Leu Val Gly Asp Lys Arg
 20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 439 base pairs
   (B) TYPE: NUCLEIC ACID
   (C) STRANDEDNESS: DOUBLE
   (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
   (A) ORGANISM: Homo Sapiens
   (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
   (A) NAME/KEY: sig_peptide
   (B) LOCATION: 200..361
   (C) IDENTIFICATION METHOD: Von Heijne matrix
   (D) OTHER INFORMATION: score 6
       seq IVFIFLILLNTAA/QV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

```
ATTGAAAGAT GGTAAAATGG TGCAGAAGGG GACTTACACT GAGTTCCTAA AATCTGGTAT   60

AGATTTTGGC TCCCTTTTAA AGAAGGATAA TGAGGAAAGT GAACAACCTC CAGTTCCAGG  120

AACTCCCACA MYAAGGGAAT CGTACCCTTC TCAGAGTCTT CGGTTTGGTC TCAACAATCT  180

TCTAGACCCT CCTTGAAAG ATG GTG CTC TGG AGA GCC AAG ATA CAN MGG AAT   232
                     Met Val Leu Trp Arg Ala Lys Ile Xaa Arg Asn
                                     -50                 -45

GTC CCA GTT ACA CTA TCA GAG GAG AAC CGT TCT GAA GGA AAA GTT GGT    280
Val Pro Val Thr Leu Ser Glu Glu Asn Arg Ser Glu Gly Lys Val Gly
            -40                 -35                 -30

TTT CAG GCC TAT AAG AAT TAC TTC AGA GCT GGT GCT CAC TGG ATT GTC    328
Phe Gln Ala Tyr Lys Asn Tyr Phe Arg Ala Gly Ala His Trp Ile Val
        -25                 -20                 -15
```

```
TTC ATT TTC CTT ATT CTC CTA AAC ACT GCA GCT CAG GTT GCC TAT GTG    376
Phe Ile Phe Leu Ile Leu Leu Asn Thr Ala Ala Gln Val Ala Tyr Val
    -10                 -5                  1               5

CTT CAA GAT TGG TGG CTT TCA TAC TGG GCA AAC AAA CAA AGT ATG CTA    424
Leu Gln Asp Trp Trp Leu Ser Tyr Trp Ala Asn Lys Gln Ser Met Leu
                10                  15                  20

AAT GTC ACT GTA AAT                                                439
Asn Val Thr Val Asn
            25
```

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 232 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 125..178
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6
seq FTSVLWLTSPSQP/NT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

```
ATGTAGTGAA TAAAGTTTGA GAACCACTGA CTTGAACTTT AGCATGATTT GATACACAGG   60

GTCCTCTGTA ATCGTACTTC GTTCTGCTTT AAGGCTGTTG GGCTGTCTCC TCCAACCCAT  120

CCKK ATG TTG TTG TAK TTT TTC ACC TCK GTC CTT TGG CTT ACG TCA CCN   169
     Met Leu Leu Xaa Phe Phe Thr Ser Val Leu Trp Leu Thr Ser Pro
                 -15                 -10                 -5

TCC CAA CCT AAT ACC TGC CCT TCT AGT CTT CTG TGT ACT TAT CCA AAT    217
Ser Gln Pro Asn Thr Cys Pro Ser Ser Leu Leu Cys Thr Tyr Pro Asn
            1               5                   10

CTA AAC CCT CCA TGG                                                232
Leu Asn Pro Pro Trp
    15
```

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 229 base pairs

(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 140..205
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.9
seq IILGCLALFLLLQ/RK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

```
AACAGTTACG AAGGAGAGCT GCAAAAGTTG CAGCAGAAAG GTTGGGAGTC CCGACAGGTT   60

CCGTAGCCCA CAGAAAAGAA GCAAGGGACG GCAGGACTGT TTCACACTTT TCTGCTTCTG  120

GAAGGTGCTG GACAAAAAC ATG GAA CTA ATT TCC CCA ACA GTG ATT ATA ATC   172
                    Met Glu Leu Ile Ser Pro Thr Val Ile Ile Ile
                            -20                 -15

CTG GGT TGC CTT GCT CTG TTC TTA CTC CTT CAG CGG AAG AAT TTG CGC    220
Leu Gly Cys Leu Ala Leu Phe Leu Leu Leu Gln Arg Lys Asn Leu Arg
    -10                 -5                   1               5

AGA CCC TGG
Arg Pro Trp                                                        229
```

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 292 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 134..274
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.9
seq TWLGLLSFQNLHC/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

```
ATCATTTTCT TATCCCTGCT GATTTCAAAC CTTCCCATGG TTTAGAAGCA TAACCTGTAA   60

TGTAATGCAA GTCCCCTAAC TCCCTGGTTG CTAACATTAA CTTCCTTAAG TAATAATCAA  120

TGAAAGAVAT TCT ATG CAT GGT TTT GAA ATA ATA TCC TTG AAA GAG GAA     169
               Met His Gly Phe Glu Ile Ile Ser Leu Lys Glu Glu
                       -45                 -40

TCA CCA TTA GGA AAG GTG AGT CAG GGT CCT TTG TTT AAT GTG ACT AGT    217
Ser Pro Leu Gly Lys Val Ser Gln Gly Pro Leu Phe Asn Val Thr Ser
-35                 -30                 -25                 -20

GGC TCA TCA TCA CCA GTG ACC TGG TTG GGC CTA CTC TCC TTC CAG AAC    265
Gly Ser Ser Ser Pro Val Thr Trp Leu Gly Leu Leu Ser Phe Gln Asn
                -15                 -10                  -5

CTG CAT TGC TTC CCA GAC CTC CCC GGG                                292
Leu His Cys Phe Pro Asp Leu Pro Gly
              1               5
```

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 270..437
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.9
seq NTLFLHLSGLSAA/DT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

```
AAGCTCTGAG ACAGGAGCCC AGCCCTGGGA TTTTCAGGTG TTTTCATTTG GTGGTCAGGC   60

CTGAACAGAG TGTTTTCCTT TGGTGGTCAG GACTGAGCAG AGAGACCTCA CCATGGAGCT  120

TKGGSYGKTG CKGGCTTTTT CTTGTGGCCA TTTTGAAAGA TGTCCGGTCT GAGGGACAAC  180

TATTGGAATC TGGGGGAAGT TCGGTCCAGC CCGGGGAGTC CCTGCGACTC TCCTGTGCAG  240

CCGCTGGATT CGCNTTTCGC AATTTTGCC ATG ACT TGG GTC CGC CAC GCT CCA    293
                                Met Thr Trp Val Arg His Ala Pro
                                    -55                 -50

GGG AAG AGT CTG GAA TGG GTC GCA ACC GTC ACA GAT GGT GGT GAT AAG    341
Gly Lys Ser Leu Glu Trp Val Ala Thr Val Thr Asp Gly Gly Asp Lys
        -45                 -40                 -35
```

```
ACC TTT TAT GCG GCC TCC GTG AAG GGC CGC TTC AAC GTC TCC AGG GAC     389
Thr Phe Tyr Ala Ala Ser Val Lys Gly Arg Phe Asn Val Ser Arg Asp
        -30                 -25                 -20

AAT TCC AAG AAC ACG TTA TTT CTG CAT TTG AGC GGC CTG AGT GCC GCC     437
Asn Ser Lys Asn Thr Leu Phe Leu His Leu Ser Gly Leu Ser Ala Ala
    -15                 -10                 -5

GAC ACG GGC TGG TGG GGG ATC                                         458
Asp Thr Gly Trp Trp Gly Ile
  1             5
```

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 226 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 143..184
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.8
seq LTSFFSLTANCQS/AG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

```
AACATACCCT TCAGGTTTAG GTCTTTCTTA GGTAAAGTTT TAACTTTAGT ATATCTTCCT    60

CAGGGCGGCC TTCTCCTTCC CCCTAGTAAG TGRAGAAACC CTTGTGTKTC TGCCCTCTGA   120

ACTCACCGCA TTTGGGATTA CC ATG CTA ACA TCC TTT TTT TCA CTG ACT GCA    172
                         Met Leu Thr Ser Phe Phe Ser Leu Thr Ala
                                         -10                 -5

AAT TGC CAG AGT GCA GGA ACT ATC TCA TTT GCT GCT TTC TCC CTA ATG     220
Asn Cys Gln Ser Ala Gly Thr Ile Ser Phe Ala Ala Phe Ser Leu Met
                 1                   5                  10

CCT GGA                                                             226
Pro Gly
```

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 140 base pairs
(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 72..125
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.8
seq LTPLFFMXPTGFS/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

```
ACTTCCCTTC CCCCTCTAGC ATTGCTACCT TCTCTCCTAC ACGCACGCAG GCATATAAAC   60

GTAGGTTTTT G ATG CTC CTC TGC CTG TTG ACC CCG CTA TTT TTC ATG TTK   110
             Met Leu Leu Cys Leu Leu Thr Pro Leu Phe Phe Met Xaa
                         -15                 -10

CCA ACA GGT TTT TCT TCC CCC AGT CCT GGG                             140
Pro Thr Gly Phe Ser Ser Pro Ser Pro Gly
 -5                  1               5
```

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 288 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 178..240
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.7
seq HSLFLSLLGLCPS/KT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

```
AATTGGCGCG GGGCGTCCGT AGCCACGGCA ACAGGTTGCT TCTGCAGTCT GAGCTGAGCG   60

CCTTTCGCAC GACTTGGAGT TACGGTTTAT TTGATACCCC GGTACCCCTA CGCAAGCAAG  120

CCCACATCGA CACACATTCA CACACGCCCT TCAGCACCCC CTCCCAGCAC CACGACC     177
```

```
ATG GAC GAC GAC TAT GAA GCG TAC CAC AGT CTG TTC TTG TCG CTG CTC     225
Met Asp Asp Asp Tyr Glu Ala Tyr His Ser Leu Phe Leu Ser Leu Leu
    -20                 -15                 -10

GGA CTC TGC CCG TCT AAG ACT CCC ATC AAT GAA AAT GCT CCC GTC TTT     273
Gly Leu Cys Pro Ser Lys Thr Pro Ile Asn Glu Asn Ala Pro Val Phe
 -5                   1               5                  10

GAT CCT GAA CCG GTC                                                 288
Asp Pro Glu Pro Val
                 15
```

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 393 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 298..354
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.7
seq WLVWLLLGHMVVS/QM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

```
CTCTTGCCTC AGGCTTGGAG GCCTCCGAGC AGCAACATCG TCCCAATTAT ACCCCGTTGG    60

AGCATCTTCA GATCTTCCAC TCTTTTCACA ACGCAATCAA AATCTTCGTA CCCATTTTGC   120

AGTAGTGATC TCTAAACTCT CAGCGTAGGC ATCGGGAACC TTCGTGCCAA GGAGCCATGC   180

TGCCCCGATG GGAACTGGCA CTTTACCTAC TTGCCTCACT AGGCTTCCAC TTCTATTCCT   240

TCTATTAAGT TTACAAAGTC TCCAGAGGAT GCGACCGACT TTGAGTGGAG CTTCTGG      297

ATG GAA TGG GGG AAG CAG TGG CTG GTG TGG CTT CTC CTT GGC CAC ATG     345
Met Glu Trp Gly Lys Gln Trp Leu Val Trp Leu Leu Leu Gly His Met
                -15                 -10                  -5

GTA GTG TCT CAA ATG GCC ACA CTG CTG GCA AGA AAG CAC AGA CCC TGG     393
Val Val Ser Gln Met Ala Thr Leu Leu Ala Arg Lys His Arg Pro Trp
             1                   5                  10
```

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 281 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 135..251
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.7
seq LTQGVLWILVIQA/VP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

```
ATATACAGAG AATAAACGTC ATCCCTCTAA CATTAATATG TTCAGTTTTA TGTACCTGAG   60

AGTTGATGGT TTAATTTGTG GGTTTGCCCA GACTCTCTTG CGACTTCTCT CATCATCTGC  120

TCTTTAGCAC TTCC ATG AGA CGG GGC AAG AGA TTG TTG GAG TCT CAA TCC    170
                Met Arg Arg Gly Lys Arg Leu Leu Glu Ser Gln Ser
                                -35                 -30

AGC AGC CCG AAA GCC TGT CTG CAG CTT GGG TTT GAG ACT GAA CTA ACT    218
Ser Ser Pro Lys Ala Cys Leu Gln Leu Gly Phe Glu Thr Glu Leu Thr
        -25                 -20                 -15

CAG GGT GTT TTG TGG ATT TTA GTT ATC CAG GCT GTC CCT GTT CCC TCA    266
Gln Gly Val Leu Trp Ile Leu Val Ile Gln Ala Val Pro Val Pro Ser
    -10                 -5                   1               5

TTA ACA AAA ACA AAA                                                281
Leu Thr Lys Thr Lys
                10
```

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 276 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 205..264
(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5.7
seq ALLESVVWLPCHG/RG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

```
AGACCAGGCC CATTTCTCAG AAGCCTTTGG CTCCCCTGAG ATGCCAAATA GCCGCTCACT   60

CTTCCGCCTC CACGGACTGG CTTTGGTGTT CATGCTGGTT GGGATGTCTA CTATGGACCT  120

GCTGAGCACA GGGCTGGGTT CCTGGGGCAC AGAGTTGATG CTTATGGCCC AGGAACTGCT  180

GGGCCCCAGG ACTGGGCGGT TTCC ATG GTT GCT GCC ACA GAA GCA GCA TTG     231
                           Met Val Ala Ala Thr Glu Ala Ala Leu
                           -20                 -15

CTG GAG TCA GTA GTG TGG CTG CCT TGC CAT GGC CGT GGT GGG TCT        276
Leu Glu Ser Val Val Trp Leu Pro Cys His Gly Arg Gly Gly Ser
    -10                 -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 356..412
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.6
seq VSLPLLSSWGSTA/WT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

```
AATTACAGCT CTACAATGCA CCAGACGGAC CCATCTGGAT TCTTTCGGGG CTCTTAGCCC   60

TAGAAATAGC ATCATTTCTT CAAACTGGTG AGTCCTCCTG TCTAAAATCA GGATGCAGAG  120

AGTTGATGCA CGGCATGGCA CAGGATGCTG GGCAAGGCTG GCAGGCCCGG GAGAGCCTGT  180

GGCCAGCCTG GGTCCAGGAA GTGGGCAGCT GCCACAGAGG GGCCTCCGAG GCTAGCTGCC  240

TCCTAACTTC CTCACGGCAC ACCATTCTGC CGTCCTGAGT CTTCTCAAGG TTGGAAGGTG  300

CCCAGATCCA GGGAGATGGT GCTGGCTCTT TGGTGGCTGT GGAGTGTCCA GACAG ATG  358
                                                             Met

AGC TGG AAT CCT TCA GTT TCT CTG CCT CTC CTG TCA AGT TGG GGT AGC    406
Ser Trp Asn Pro Ser Val Ser Leu Pro Leu Leu Ser Ser Trp Gly Ser
```

```
        -15                 -10                 -5

ACT GCT TGG ACT CTT                                                421
Thr Ala Trp Thr Leu
          1
```

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 193 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 53..118
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.6
seq LILLSLHLERRWT/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

```
ACAATAATAA CTAATGAGAT TAAAATTTAA AACAGGTGTC TGATAATCCT TG ATG AAG   58
                                                          Met Lys

AGA ATT CAG GGG ATA TTG TTC CTG ATT TTG CTT TCT CTC CAC TTG GAA    106
Arg Ile Gln Gly Ile Leu Phe Leu Ile Leu Leu Ser Leu His Leu Glu
-20                 -15                 -10                 -5

AGG AGG TGG ACG AGC CCA TCA GAC CAC AGC CTG TTG CTA GGA GGA AAT    154
Arg Arg Trp Thr Ser Pro Ser Asp His Ser Leu Leu Leu Gly Gly Asn
                  1                 5                  10

TCC TTG GCT CAA CAT GCA GAA AGT GTA GTA CGC CAA GGG                193
Ser Leu Ala Gln His Ala Glu Ser Val Val Arg Gln Gly
        15                  20                  25
```

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 435 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 298..402
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.5
seq LLTFGLEVCLAAG/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

```
AAAGGAAGGG GGGGCGGAAC CAGCCTGCAC GCGCTGGCTC CGGGTGACAG CCGCGCGCCT    60

CGGCCAGGAT CTGAGTGATG AGACGTGTCC CCACTGAGGT GCCCCACAGC AGCAGGTGTT   120

GAGCATGGGC TGAGAAGCTG GACCGGCACC AAAGGGCTGG CAGAAATDVG CGCCTGGCTG   180

ATTCCTAGGC AGTTGGCRGC AGCAAGGAGG AGAGGCCGCA GCTTCTGGAG CAGAGCCGAG   240

ACGAAGCAGT TCTGGAGTGC CTGAACGGCC CCCTGAGCCC TACCCGCCTG GCCCACT      297

ATG GTC CAG AGG CTG TGG GTG AGC CGC CTG CTG CGG CAC CGG AAA GCC     345
Met Val Gln Arg Leu Trp Val Ser Arg Leu Leu Arg His Arg Lys Ala
-35                 -30                 -25                 -20

CAG CTC KKG CTG GKC AAC CTG CTA ACC TTT GGC CTG GAG GTG TGT TTG     393
Gln Leu Xaa Leu Xaa Asn Leu Leu Thr Phe Gly Leu Glu Val Cys Leu
            -15                 -10                  -5

GCC GCA GGA TCA CCT ATG TGC CGC CTC TGC TGC TGG AAG TGG             435
Ala Ala Gly Ser Pro Met Cys Arg Leu Cys Cys Trp Lys Trp
              1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 27..80
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.5
seq PFALVTSCSSVFS/GD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

```
AGACGGAACC AGCTCGCGAG CGCACT ATG GCT GCT GGC GTA CCC TTT GCG TTA     53
```

```
                                    Met Ala Ala Gly Val Pro Phe Ala Leu
                                                -15                 -10

GTC ACC AGC TGC TCC TCC GTC TTC TCA GGA GAC CAG CTG GTC CAA CAT     101
Val Thr Ser Cys Ser Ser Val Phe Ser Gly Asp Gln Leu Val Gln His
                 -5                   1               5

ATC CTT GGA ACA GAA GAT CTT ATT GTG GAA GTG ACT TCT AAT GAT GCT     149
Ile Leu Gly Thr Glu Asp Leu Ile Val Glu Val Thr Ser Asn Asp Ala
         10                  15                  20

GTG AGA TTT TAT CCC TGG ACC ATT GAT AAT AAA TAC TAT TCA GCA GAC     197
Val Arg Phe Tyr Pro Trp Thr Ile Asp Asn Lys Tyr Tyr Ser Ala Asp
     25                  30                  35

ATC AAT CTA TGT GTG GTG CCW AAC AAA TTT CTT GTT ACT GCA GAG ATT     245
Ile Asn Leu Cys Val Val Pro Asn Lys Phe Leu Val Thr Ala Glu Ile
 40                  45                  50                  55

GCA GAA TCT GTC CAA GCA TTT GTG GTT TAC TTT GAC DKC ACA CAA RAA     293
Ala Glu Ser Val Gln Ala Phe Val Val Tyr Phe Asp Xaa Thr Gln Xaa
                 60                  65                  70

TCG GGC CTT GAT AGT GTC TCC TCA TGG CTT CCA CTG GCA AAA GCA TGG     341
Ser Gly Leu Asp Ser Val Ser Ser Trp Leu Pro Leu Ala Lys Ala Trp
             75                  80                  85

TTA CCT GAG GTG ATG ATC TTG GTC TGC GAT AGA GTG TCT GAA GAT GGT     389
Leu Pro Glu Val Met Ile Leu Val Cys Asp Arg Val Ser Glu Asp Gly
         90                  95                 100

ATA                                                                 392
Ile
```

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 358 base pairs
    (B) TYPE: NUCLEIC ACID
    (C) STRANDEDNESS: DOUBLE
    (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Homo Sapiens
    (F) TISSUE TYPE: Prostate

(ix) FEATURE:
    (A) NAME/KEY: sig_peptide
    (B) LOCATION: 290..331
    (C) IDENTIFICATION METHOD: Von Heijne matrix
    (D) OTHER INFORMATION: score 5.5
        seq TVFLXFCFPRCHS/DS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

```
AATGCTGAAC GGWCACACAG AGAGGAAACW ATAAATKKYA GCTACTATGC AATAAATATC    60
```

```
TCAAGTTTTA ACGAAGAAAA ACATCATTGC AGTGAAATAA AAAATTTTAA AATTTTAGAA  120

CAAAGCTAAC AAATGGCTAG TTTTCTATGN TTCTTCTTCA AACGCTTTCT TTGAGGGRGM  180

AAGAGTCAMA CAAACAAGCA GTTTTACCTA AAATAAAGAA CTAGTTTTAG AGGTCAGAMG  240

AMAGGMGCAA GTTTTGCGAG WGGCACGGAA GGAGTGTGCT GGCAGTACA ATG ACA GTT  298
                                                      Met Thr Val

TTC CTT TMN TTT TGC TTT CCT CGC TGC CAT TCT GAC TCA CAT ARG RTG  346
Phe Leu Xaa Phe Cys Phe Pro Arg Cys His Ser Asp Ser His Xaa Xaa
    -10                 -5                  1                   5

CAG CAA TCA GCG                                                   358
Gln Gln Ser Ala
```

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 44..187
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.4
seq ILLEVFVWNGLQG/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

```
AASTTCTTCC TGCCAAGAGA ACAATGCCGA GAAACAGAGC GAA ATG KTT CCA AAT    55
                                                Met Xaa Pro Asn
                                                            -45

AAT TTT TGG CAA AAA CTT GGA AGA AAA AAA CCC CGC ATA TTT ACC TGT   103
Asn Phe Trp Gln Lys Leu Gly Arg Lys Lys Pro Arg Ile Phe Thr Cys
                -40                 -35                 -30

ACC CAG AGC TCC ACA GGT GAG GCG GCA GTT AAA GCA GAA AAT CTA ATT   151
Thr Gln Ser Ser Thr Gly Glu Ala Ala Val Lys Ala Glu Asn Leu Ile
            -25                 -20                 -15

CTT CTG GAA GTT TTT GTC TGG AAC GGA CTC CAG GGT CTT CCT TCG GAG   199
Leu Leu Glu Val Phe Val Trp Asn Gly Leu Gln Gly Leu Pro Ser Glu
        -10                 -5                  1

CTG TCA GAT ACA AGT GGA TCC TCT AAG AAA CTT GGG AGC CTT GTG GGC   247
Leu Ser Asp Thr Ser Gly Ser Ser Lys Lys Leu Gly Ser Leu Val Gly
```

```
 5                   10                  15                  20

TGG TGG AGA ACT CTC AAG ATG GCA CCA GCC TGT CTA TGG TCT ATG TGG     295
Trp Trp Arg Thr Leu Lys Met Ala Pro Ala Cys Leu Trp Ser Met Trp
                 25                  30                  35

GAA TCA CCG CCA CGG                                                 310
Glu Ser Pro Pro Arg
             40
```

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 284 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 66..173
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.3
seq ALYIMCVPHSVWG/CA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

```
AAGTCCAGAG GCCTGGCCCT GCCAAGAAGG CGCTCTCCGG AATCAACACC TGGGGGCTTG    60

GAAGG ATG TTT CGC TCA GAT CGA ATG TGG ARC TGC CAT TGG AAA TGG AAG  110
      Met Phe Arg Ser Asp Arg Met Trp Xaa Cys His Trp Lys Trp Lys
          -35                 -30                 -25

CCC AGT CCT CTC CTG TTC TTA TTT GCT TTA TAT ATC ATG TGT GTT CCT     158
Pro Ser Pro Leu Leu Phe Leu Phe Ala Leu Tyr Ile Met Cys Val Pro
    -20                 -15                 -10

CAC TCA GTG TGG GGA TGT GCC AAC TGC CGA GTG GTT TTG TCC AAC CCT     206
His Ser Val Trp Gly Cys Ala Asn Cys Arg Val Val Leu Ser Asn Pro
 -5                   1               5                  10

TCT GGG ACC TTT ACT TCT CCA TGC TAC CCT AAC GAC TAC CCA AAC AGC     254
Ser Gly Thr Phe Thr Ser Pro Cys Tyr Pro Asn Asp Tyr Pro Asn Ser
             15                  20                  25

CAG GCT TGC ATG TGG ACG CTC CGA GAC CCC                             284
Gln Ala Cys Met Trp Thr Leu Arg Asp Pro
         30                  35
```

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 398 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 123..215
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 5.3
 seq LVALSSELPFLGA/GV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

```
TCCTTCATCT TGTGTTCTAA AACCTTGCAA GTTCAGGAAG AAACCATCTG CATCCATATT    60

GAAAACCTGA CACAATGTAT GCAGCAGGCT CAGTGTGAGT GAACTGGAGG CTTCTCTACA   120

AC ATG ACC CAA AGG AGC ATT GCA GGT CCT ATT TGC AAC CTG AAG TTT      167
   Met Thr Gln Arg Ser Ile Ala Gly Pro Ile Cys Asn Leu Lys Phe
       -30                 -25                 -20

GTG ACT CTC CTG GTT GCC TTA AGT TCA GAA CTC CCA TTC CTG GGA GCT     215
Val Thr Leu Leu Val Ala Leu Ser Ser Glu Leu Pro Phe Leu Gly Ala
    -15                 -10                  -5

GGA GTA CAG CTT CAA GAC AAT GGG TAT AAT GGA TTG CTC ATT GCA ATT     263
Gly Val Gln Leu Gln Asp Asn Gly Tyr Asn Gly Leu Leu Ile Ala Ile
  1               5                  10                  15

AAT CCT CAG GTA CCT GAG AAT CAG AAC CTC ATC TCA AAC ATT AAG GAA     311
Asn Pro Gln Val Pro Glu Asn Gln Asn Leu Ile Ser Asn Ile Lys Glu
             20                  25                  30

ATG ATA ACT GAA GCT TCA TTT TAC CTA TTT AAT GCT ACC AAG AGA AGA     359
Met Ile Thr Glu Ala Ser Phe Tyr Leu Phe Asn Ala Thr Lys Arg Arg
         35                  40                  45

GTA TTT TTC AGA AAT ATA AAG ATT TTA ATA CCT GCC CAG                 398
Val Phe Phe Arg Asn Ile Lys Ile Leu Ile Pro Ala Gln
     50                  55                  60
```

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 324 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 187..228
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.3
seq IIPLLLLLRSACN/VH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```
ACTCCAGGAG CCGGGACCAA AATAACCGGG CGGGAGGGGA CACCTCGCAG AGATGGATCT    60

CGAACTCCTG GGCTCAAGCG ATCCTTTCAC CTTGGCCTCT CAAGTAGCTG GGACCACATT   120

TGCTCACCAG CTGGCCCAAG ACCAGACTGG GCAACATGGG TCATCCTCCT CTAAGATTCC   180

AGGACC ATG ATC ATC CCT CTA TTG CTA CTT CTT AGA TCA GCT TGT AAT      228
       Met Ile Ile Pro Leu Leu Leu Leu Leu Arg Ser Ala Cys Asn
                       -10                 -5

GTC CAT CTC CCC CAC CAG ACT GCG TCT CCA GCA TCT CTG AGT CCC CAG     276
Val His Leu Pro His Gln Thr Ala Ser Pro Ala Ser Leu Ser Pro Gln
  1               5                  10                  15

GGC CTG GCC TGG GGC TTG CTA CAT GGT GGG TGC TCA GTA ACT GTG AGA     324
Gly Leu Ala Trp Gly Leu Leu His Gly Gly Cys Ser Val Thr Val Arg
             20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 231..287
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.3
seq VLLLSXNLNLIIQ/SS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
ACTATGGAGG GAAGACACAG GGAAAGGAGT ATTTCAAAAA CTTTAAATAA TTGTACATAA    60
```

```
TTGGAGCAAG TGAGAAGACA AGTKAGAGGT AAGCWGKTRT TGAGAATAGG GGKCTGATTG  120

TGCCAGCTTT GTATACVATT ATNAGGAACN DGGACTTTGT CCTGAAGGTA ACTGGGCAAT  180

TGTTGAGGTC ACCACCATCT ACTGTCTGGA TTACCGAGGA AACTTTCTAA ATG TMS     236
                                                       Met Xaa

TCT CCA CTT CCA GTC CTG CTC CTC TCA TKC AAT CTC AAC CTA ATA ATT    284
Ser Pro Leu Pro Val Leu Leu Leu Ser Xaa Asn Leu Asn Leu Ile Ile
        -15                 -10                     -5

CAG AGT AGT
Gln Ser Ser                                                        293
    1
```

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 244..381
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.2
seq LLTFLVFTXKLSS/LN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
ACACTGAAAT CAATCTGTTC AATAGCATTA TACCATATTT GACATACCAT AGCCATGTTA   60

ATCTGATATT GTAGAATAGC ATAGTAKAAT AATAATAACT CCTAACTCAA GGATGTTGWG  120

WKCCTTTATA ACCAGCAATC CATGTTARAT ATTAGCACAG TGCCTAAAAC ATATTAAGCA  180

TTCAATAAAT GATCGCTACT ATTTTTACTA ACATCCTACA GATTTGGAAA TTGAGTCTTA  240

GAA ATG TTA ATG TGT AAA ATG CTA AAG AGC CAA AAA AAC TGC CAG GAA    288
    Met Leu Met Cys Lys Met Leu Lys Ser Gln Lys Asn Cys Gln Glu
        -45                 -40                     -35

AAT ATR ARA ATT AAA ATC ATT TTA TTT CTG AAA CCC ATG TGT TCC CCC    336
Asn Xaa Xaa Ile Lys Ile Ile Leu Phe Leu Lys Pro Met Cys Ser Pro
    -30                 -25                     -20

CAA TAT CTT CTA ACA TTT CTA GTA TTT ACA GRA AAA CTT TCA AGT CTC    384
Gln Tyr Leu Leu Thr Phe Leu Val Phe Thr Xaa Lys Leu Ser Ser Leu
-15                 -10                     -5                  1
```

```
AAT ATC RGA AAG TTT CAT                                    402
Asn Ile Xaa Lys Phe His
              5
```

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 306..461
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.2
seq IIVILHCAASIIS/CP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```
AAGTATTAAA TTTAAAAAGA TAAATCTGCC CTATTCTAAT CATGTCTTTG TCTTCTGTTT   60

ATTCAAGTGT ATTCCATTTG CTTTCGGGAA TATTTGGATG TTTTAGAACT AACATTCTGC  120

TTTAATAATC CAAACACRCK AYMAKTYCCA TCAATTTGAG TCTCTTAAAA TGTTACACTG  180

AAATGAATCT CTCTGAAGAT GGACTTATTG ATTTCTATAT TCTTCCTCTA GCATCATGAA  240

ATTTGACCTC TTCAGCCGTG CATGGTTAAC ACTCAGATAA CCCATCTCCT TGAGAAGAAC  300

CCCTG ATG AAR AAG AAA TCC TCT CCA AAT CAA TAT CTT CAT TCA TCA CTA  350
      Met Lys Lys Lys Ser Ser Pro Asn Gln Tyr Leu His Ser Ser Leu
              -50                 -45                 -40

CAC TRS ATA CGN CTA TTT TCC TTC CTC CAT TTC TCA GAG GAA GGA GTT    398
His Xaa Ile Arg Leu Phe Ser Phe Leu His Phe Ser Glu Glu Gly Val
        -35                 -30                 -25

CTA TTA CTT GCC ATT GAT CTT AAA ATT ATA GTT ATC CTC CAC TGT GCT    446
Leu Leu Leu Ala Ile Asp Leu Lys Ile Ile Val Ile Leu His Cys Ala
    -20                 -15                 -10

GCA TCC ATA ATT TCA TGT CCC TCA                                    470
Ala Ser Ile Ile Ser Cys Pro Ser
 -5                   1
```

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 334 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 116..184
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.1
seq ATSVSLEAQSCFA/WP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

```
ATTTTTGAAA ACTGTAATGC TTTAAAACTT ACTTTATTGG ATCTCTTTGC AGCTTTTGAC    60

ACAGTGAACC ACTTTCCTTT CCTGAAATGC TTTCCTCTCT TGGCTTTCTG ATGCC ATG   118
                                                             Met

TTC TCC TGT TTC TTC TCT ACT TCT CTG GCC ACT TCT GTC TCC TTA GAA     166
Phe Ser Cys Phe Phe Ser Thr Ser Leu Ala Thr Ser Val Ser Leu Glu
        -20                 -15                 -10

GCT CAG TCT TGC TTT GCC TGG CCC TTG ATT GTT AGT TTT CCC CAG GGC     214
Ala Gln Ser Cys Phe Ala Trp Pro Leu Ile Val Ser Phe Pro Gln Gly
     -5                   1               5                  10

TCA CTT CTT AGC CCC TTT CTC CTC ATG TCT TAT AAT TTG AGT CAT CTC     262
Ser Leu Leu Ser Pro Phe Leu Leu Met Ser Tyr Asn Leu Ser His Leu
                 15                  20                  25

ATC TAC TCT GGG GAG TTG AAT GGT CGC TTG TAT GCT GAA AAC TCC CAA     310
Ile Tyr Ser Gly Glu Leu Asn Gly Arg Leu Tyr Ala Glu Asn Ser Gln
             30                  35                  40

ATT TGT ATC TGT AGC CCA GCC GGG                                     334
Ile Cys Ile Cys Ser Pro Ala Gly
        45                  50
```

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 302 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 78..227
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.1
seq RTALILAVCCGSA/SI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

```
AGTTTCCAAG GGAAGGAGCA GCGTGTGGGA AAGCACAGAA GAGTGAGAAG GAAGCGACTA    60

AATTTTATTT ACTTTCT ATG CAT CAT GGC CTC ACA CCA CTG TTA CTT GGT      110
                   Met His His Gly Leu Thr Pro Leu Leu Leu Gly
                   -50                 -45                 -40

GTA CAT GAG CAA AAA CAG CAA GTG GTG AAA TTT TTA ATC AAG AAA AAA     158
Val His Glu Gln Lys Gln Gln Val Val Lys Phe Leu Ile Lys Lys Lys
                -35                 -30                 -25

GCA AAT TTA AAT GCA CTG GAT AGA TAT GGA AGA ACT GCT CTC ATA CTT     206
Ala Asn Leu Asn Ala Leu Asp Arg Tyr Gly Arg Thr Ala Leu Ile Leu
            -20                 -15                 -10

GCT GTA TGT TGT GGA TCG GCA AGT ATA GTC AGC CTT CTA CTT GAG CAA     254
Ala Val Cys Cys Gly Ser Ala Ser Ile Val Ser Leu Leu Leu Glu Gln
         -5                   1               5

AAC ATT GAT GTA TCT TCT CAA GAT CTA TCT GGA CAG ACG GCC CCC GGG     302
Asn Ile Asp Val Ser Ser Gln Asp Leu Ser Gly Gln Thr Ala Pro Gly
 10                  15                  20                  25
```

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 381 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 319..369
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.1
seq IYFFACFQALTSS/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

```
AGAGTAGGCG GAGACAGAGA GGCTGTATTT CAGTGCAGCC TGCCAGACCT CTTCTGGAGG    60
```

```
AAGACTGGAC AAAGGGGGTC ACACATTCCT TCCATACGGT TGAGCCTCTA CCTGCCTGGT  120

GCTGGTCACA GTTCAGCTTC TTCATGATGG TGGATCCCAA TGGCAATGAA TCCAGTGCTA  180

CATACTTCAT CCTAATAGGC CTCCCTGGTT TAGAAGAGGC TCAGTTCTGG TTGGCCTTCC  240

CATTGTGCTC CCTCTACCTT ATTGCTGTGC TAGGTAACTT GACAATCATC TACATTGTGC  300

GGACTGAGCA CAGCCTGC ATG AGC CCA TGT ATA TAT TTC TTT GCA TGC TTT    351
                    Met Ser Pro Cys Ile Tyr Phe Phe Ala Cys Phe
                                -15                 -10

CAG GCA TTG ACA TCC TCA TCT CCA CCT CAG                            381
Gln Ala Leu Thr Ser Ser Ser Pro Pro Gln
    -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 318 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 49..141
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.1
seq VSGASGFLPPARS/RI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

```
CTTTCTGTGT CTCCTTTCCT CCGCCTCAGT TTGGGGCGGG TCGGGGGA ATG GCT GAG     57
                                                     Met Ala Glu
                                                         -30

GAG ATG GAG TCG TCG CTC GAG GCA AGS TTT TCG TCC AGC GGG GCA GTG    105
Glu Met Glu Ser Ser Leu Glu Ala Xaa Phe Ser Ser Ser Gly Ala Val
            -25                 -20                 -15

TCA GGG GCC TCA GGG TTT TTG CCT CCT GCC CGC TCC CGC ATC TTC AAG    153
Ser Gly Ala Ser Gly Phe Leu Pro Pro Ala Arg Ser Arg Ile Phe Lys
        -10                 -5                  1

ATA ATC GTG ATC GGC GAC VBC AAT GTG GGC AAG ACA TGC CTG ACC TAC    201
Ile Ile Val Ile Gly Asp Xaa Asn Val Gly Lys Thr Cys Leu Thr Tyr
 5                  10                  15                  20

CGC TTC TGC GCT GGC CGC TTC CCC GAC CGC ACC GAG GCC ACG ATA GGG    249
Arg Phe Cys Ala Gly Arg Phe Pro Asp Arg Thr Glu Ala Thr Ile Gly
```

```
              25                   30                   35

GTG GAT TTC CGA GAA CGA GCG GTG GAG ATT GAT GGG GAG CGC ATC AAG    297
Val Asp Phe Arg Glu Arg Ala Val Glu Ile Asp Gly Glu Arg Ile Lys
          40                   45                   50

ATC CAG CTA TGG GAC ACA GCA                                         318
Ile Gln Leu Trp Asp Thr Ala
        55
```

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 61..153
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.1
seq VSGASGFLPPARS/RI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

```
AAATCTCTCA GCCTTTCTGT GTCTCCTTTC CTCCGCCTCA GTTTGGGGCG GGTCGGGGGA    60

ATG GCT GAG GAG ATG GAG TCG TCG CTC GAG GCA AGC TTT TCG TCC AGC    108
Met Ala Glu Glu Met Glu Ser Ser Leu Glu Ala Ser Phe Ser Ser Ser
    -30                 -25                 -20

GGG GCA GTG TCA GGG GCC TCA GGG TTT TTG CCT CCT GCC CGC TCC CGC    156
Gly Ala Val Ser Gly Ala Ser Gly Phe Leu Pro Pro Ala Arg Ser Arg
-15                 -10                 -5                   1

ATC TTC AAG ATA ATC GTG ATC GGC GAC TCC AAT GTD VGC AAG ACA TGC    204
Ile Phe Lys Ile Ile Val Ile Gly Asp Ser Asn Val Xaa Lys Thr Cys
              5                  10                  15

CTG ACC TAC CGC TTC TGC GCT GGC CGC TTC CCC GAC CGG                243
Leu Thr Tyr Arg Phe Cys Ala Gly Arg Phe Pro Asp Arg
        20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 153..233
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5
seq HLSLILLKPLCLP/NN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

```
ACCTTTTATA AACATTTTGT TTAACTTTTA TTGTGGTAAA ATACACATAA CACTTCTCTT     60

CTTTTAGACC TGGGCTGGTA AGAAGTGCTG AAGATGTTTT TTAGAGATTT GTGGTATGAC    120

AAATTCCACT GGGGTTTCTG ASCTTCTCAG TC ATG CTT GTC TTG GGG TCA CCA      173
                                    Met Leu Val Leu Gly Ser Pro
                                                -25

CTC CTT GGC CCT CTC CTA TGG CAC CTG TCC CTC ATT CTG CTC AAG CCC      221
Leu Leu Gly Pro Leu Leu Trp His Leu Ser Leu Ile Leu Leu Lys Pro
-20                 -15                 -10                  -5

CTA TGC CTT CCC AAC AAC TTG CCT TTA GCT CTG GGC AGA TGT CTT TGC      269
Leu Cys Leu Pro Asn Asn Leu Pro Leu Ala Leu Gly Arg Cys Leu Cys
                  1               5                  10

TTG CAC TCG                                                          278
Leu His Ser
        15
```

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 301 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 56..220
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5
seq VLFMTTAVDLVIT/EV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

```
AGAAAGGTGT TTTGGTCTTC TCCTTAGTCC AGGAAAAGAT GTACGAAATA GTGAC ATG     58
                                                            Met
                                                            -55

CAC TTA TTA GAT TTG GAA TCT ATG GGC AAA AGT TCA GAT GGA AAG TCG    106
His Leu Leu Asp Leu Glu Ser Met Gly Lys Ser Ser Asp Gly Lys Ser
                -50                 -45                 -40

TAT GTT ATT ACG GGG AGC TGG AAT CCA AAA TCC CCA CAT TTT CAA GTT    154
Tyr Val Ile Thr Gly Ser Trp Asn Pro Lys Ser Pro His Phe Gln Val
            -35                 -30                 -25

GTA AAT GAA GAA ACT CCT AAA GAT AAA GTC CTG TTT ATG ACC ACA GCT    202
Val Asn Glu Glu Thr Pro Lys Asp Lys Val Leu Phe Met Thr Thr Ala
        -20                 -15                 -10

GTA GAT TTG GTA ATA ACA GAA GTA CAG GAG CCT GTT CGA TTT CTC CTG    250
Val Asp Leu Val Ile Thr Glu Val Gln Glu Pro Val Arg Phe Leu Leu
     -5                   1               5                  10

GAG ACA AAA GTC CGC GTT TGC TCA CCT AAT GAA AGA TTA TTC TGG CCC    298
Glu Thr Lys Val Arg Val Cys Ser Pro Asn Glu Arg Leu Phe Trp Pro
                15                  20                  25

GCG                                                                301
Ala
```

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 129 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..63
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.8
seq VLFVFSSIPLTFL/FQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

```
ATG GAG AAT TTG AAA GAC TTT TAT GTG TTG TTT GTA TTC TCT AGC ATT     48
Met Glu Asn Leu Lys Asp Phe Tyr Val Leu Phe Val Phe Ser Ser Ile
    -20                 -15                 -10
```

```
CCC CTT ACA TTT CTA TTT CAG AAA TTG CCT TTT GTT TGG ATT KGA GAA       96
Pro Leu Thr Phe Leu Phe Gln Lys Leu Pro Phe Val Trp Ile Xaa Glu
 -5                  1                   5                  10

GAG ACT TTG GAG ACA TGG TAT TTG AAG AGC TGG                           129
Glu Thr Leu Glu Thr Trp Tyr Leu Lys Ser Trp
             15                  20
```

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 293..346
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.8
seq LSIFSLVLPVCRM/HR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

```
ACAATTCCAG CTTATGTGTC CCTTTTATAA ACTTGTGATA CATTTTAACT GTGTATACAC    60

ATCTCTTGCC TCTATTGGTA GAGAGTATCT GSCAKGCCTA GCATGTGCTG GATGTCATAT   120

CAGATACTCA GTGTTATTTA TTGGGCTTAC AGTGATAACC AAAGCTCACA TGTTTTAGCA   180

CTCCCACTTC CATAAAGTGG AAGATGTCCC CTCTGCCTCT TCTCTCATCC CTCCTCAAAG   240

CAGCAGGAGT GACTTACCTG ATTGACCAGT TTAAGACTAT ATCTGAGCAG GC ATG CCA   298
                                                          Met Pro

CAG TAC TGT CTC AGC ATC TTC TCT CTT GTG CTG CCT GTC TGC AGG ATG       346
Gln Tyr Cys Leu Ser Ile Phe Ser Leu Val Leu Pro Val Cys Arg Met
    -15                 -10                  -5

CAC AGG                                                               352
His Arg
  1
```

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 194 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 15..143
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.8
seq LLAFGTSCSVVLY/DP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

```
GACCAGTTGG CGAC ATG GTG GCA CCC GTG CTG GAG ACT TCT CAC GTG TTT     50
                Met Val Ala Pro Val Leu Glu Thr Ser His Val Phe
                            -40                 -35

TGC TGC CCA AAC CGG GTG CGG GGA GTC CTG AAC TGG AGC TCT GGG CCC     98
Cys Cys Pro Asn Arg Val Arg Gly Val Leu Asn Trp Ser Ser Gly Pro
    -30                 -25                 -20

AGA GGA CTT CTG GCC TTT GGC ACG TCC TGC TCC GTG GTG CTC TAT GAC    146
Arg Gly Leu Leu Ala Phe Gly Thr Ser Cys Ser Val Val Leu Tyr Asp
-15                 -10                  -5                   1

CCC CTG GGT TGT TGT TAC CAA CTT GAA TGG TCA CAC CGC CCG TTC CGG    194
Pro Leu Gly Cys Cys Tyr Gln Leu Glu Trp Ser His Arg Pro Phe Arg
              5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 134..247
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.8
seq LSWLITWFGHXLS/DF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

```
ATTGTGGTCA CATTTCTGCT GGTGGTAGGC GAGAGGCTGG CAACATCCCT GGTAGAAAAA     60
```

```
TTATCTACCC ACCACCTCAG GGATTTTATG GATCCAVCAA TGGRACAACA CCAMGCATAT  120

ATTAAACTAT CTG ATG CCC ATC ATT GAC CAG GTG AAT CCA GAG CTC CAT     169
               Met Pro Ile Ile Asp Gln Val Asn Pro Glu Leu His
                           -35                     -30

GAC TTC ATG CAG AGT GCT GAG GTA GGG ACC ATC TTT GCC CTC AGC TGG    217
Asp Phe Met Gln Ser Ala Glu Val Gly Thr Ile Phe Ala Leu Ser Trp
    -25                 -20                     -15

CTC ATC ACC TGG TTT GGG CAT GWM CTG TCT GAC TTC AGG CAC GTC GTG    265
Leu Ile Thr Trp Phe Gly His Xaa Leu Ser Asp Phe Arg His Val Val
-10                 -5                      1               5

CGG TTA TAT GAC TTC TTC CTR GCC TGC CAC CCA CTG ATG CCG ATT TAC    313
Arg Leu Tyr Asp Phe Phe Leu Ala Cys His Pro Leu Met Pro Ile Tyr
            10                  15                  20

TTT GCA GCC GTG ATT GTG TTG TAT CGC GAG CAG                        346
Phe Ala Ala Val Ile Val Leu Tyr Arg Glu Gln
        25                  30
```

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 63..209
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.7
seq GLCVLVPCSXSXX/WR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

```
AAKTKKKKGG AGCATTTCCT TCCCTGACAG CCGGACCTGG KACTGGGCTG GGGCCCTGGC  60

GG ATG GAG ACA TKC TGC CCC TGC TGC TGC TGC CCC TGC KGT GGG GDN     107
   Met Glu Thr Xaa Cys Pro Cys Cys Cys Cys Pro Cys Xaa Gly Xaa
                   -45                 -40                 -35

GGG TCC CTG CAK GAK AAG CCA GTK TAC GAG CTG CAA GTG CAG AAG TCG    155
Gly Ser Leu Xaa Xaa Lys Pro Val Tyr Glu Leu Gln Val Gln Lys Ser
            -30                 -25                 -20

GTG ACG GTG CAG GAG GGC CTG TGC GTC CTT GTG CCC TGC TCC TKC TCT    203
Val Thr Val Gln Glu Gly Leu Cys Val Leu Val Pro Cys Ser Xaa Ser
```

```
           -15                 -10                  -5

TAS SCC TGG AGA TCC TGG TAT TCC TCT CCC CCA CTC TAC GTC TAC TGG     251
Xaa Xaa Trp Arg Ser Trp Tyr Ser Ser Pro Pro Leu Tyr Val Tyr Trp
         1               5                   10

TTC CGG GAC GGG GAG ATC CCA TAC TAC GCT GAG GTT GTG GCC ACA AAC     299
Phe Arg Asp Gly Glu Ile Pro Tyr Tyr Ala Glu Val Val Ala Thr Asn
 15                  20                  25                  30

AAC CCA GAC AGA AGA KTG AAG SMD KAK AYY CAK KGG CCG ATT CCG CCT     347
Asn Pro Asp Arg Arg Xaa Lys Xaa Xaa Xaa Xaa Xaa Pro Ile Pro Pro
                 35                  40                  45

CCT TGG GGA TGT CCA GAA GAA GAA CTG                                 374
Pro Trp Gly Cys Pro Glu Glu Glu Leu
             50                  55
```

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 295..345
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.7
seq IYFFACFXXLTSS/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

```
ATTTTCAGTG CAGCCTGCCA GACCTCTTCT GGAGGAAGAC TGGACAAAGG GGGTCACACA    60

TTCCTTCCAT ACGGTTGAGC CTCTACCTGC CTGGTGCTGG TCACAGTTCA GCTTCTTCAT   120

GRWKGGTGGA TCCCAATGGC AATGAATCCA GTGCTACATA CTTCATCCTA ATAGGCCTCC   180

CTGGTTTAGA AGAGGCTCAG TTCTGGTTGG CCTTCCCATT GTGCTCCCTC TACCTTATTG   240

CTGTGCTAGG TAACTTGACA ATCATCTACA TTGTGCGGAC TGAGCACAGC CTGC ATG     297
                                                            Met

AGC CCA TGT ATA TAT TTC TTT GCA TGC TTT CAN NNA TTG ACA TCC TCA     345
Ser Pro Cys Ile Tyr Phe Phe Ala Cys Phe Xaa Xaa Leu Thr Ser Ser
    -15                 -10                  -5

TCT CCA CCT CAT CCA TGC CCA AAA TGC TGG CCA TCT TCT GGT TCA ATT     393
Ser Pro Pro His Pro Cys Pro Lys Cys Trp Pro Ser Ser Gly Ser Ile
```

```
 1               5                   10                 15

CCA CTA                                                          399
Pro Leu
```

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 110 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 12..92
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.7
seq VLKCLSFSXPSLP/GF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

```
AAGCAACCGG G ATG GGA CGG GGA GAG AGG AGG CAC TAC TGG GGA CCT AAG      50
             Met Gly Arg Gly Glu Arg Arg His Tyr Trp Gly Pro Lys
                     -25                 -20                 -15

CTG GTT CTC AAA TGC CTC TCC TTT TCS SCT CCA AGC CTC CCA GGC TTC       98
Leu Val Leu Lys Cys Leu Ser Phe Ser Xaa Pro Ser Leu Pro Gly Phe
                -10                 -5                   1

CTA TGG TCC CTA                                                      110
Leu Trp Ser Leu
            5
```

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 251 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide

(B) LOCATION: 9..164
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.7
seq LLAKALHLLKSSC/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

```
AGCCTGCG ATG TCT CAA GAT GGC GGA STG GGC GAA TTA AAG CAC ATG GTG     50
         Met Ser Gln Asp Gly Gly Xaa Gly Glu Leu Lys His Met Val
                 -50                 -45                 -40

ATG AGT TTC CGG GTG TCT GAG CTC CAG GTG CTT CTT GGC TTT GCT GGC      98
Met Ser Phe Arg Val Ser Glu Leu Gln Val Leu Leu Gly Phe Ala Gly
            -35                 -30                 -25

CGG AAC AAG AGT GGA CGG AAG CAC GAG CTC CTG GCC AAG GCT CTG CAC     146
Arg Asn Lys Ser Gly Arg Lys His Glu Leu Leu Ala Lys Ala Leu His
        -20                 -15                 -10

CTC CTG AAG TCC AGC TGT GCC CCT AGT GTC CAG ATG AAG ATC AAA GAG     194
Leu Leu Lys Ser Ser Cys Ala Pro Ser Val Gln Met Lys Ile Lys Glu
     -5                   1               5                  10

CTT TAC CGA CGA CGC TTT CCC CGG AAG ACC CTG GGG CCC TCT GAT CTC     242
Leu Tyr Arg Arg Arg Phe Pro Arg Lys Thr Leu Gly Pro Ser Asp Leu
                 15                  20                  25

TCC CTA AAG                                                         251
Ser Leu Lys
```

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 272 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 18..224
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.6
seq LGPSLSSLPSALS/LM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

```
TATTTGGCCC CAAGCCG ATG CAT CAC AGG ATG AAT GAA ATG AAC CTG AGT       50
                   Met His His Arg Met Asn Glu Met Asn Leu Ser
                                   -65                 -60
```

```
CCA GTG GGG ATG GAG CAG CTG ACT TCA TCC TCT GTG AGC AAT GCC TTG       98
Pro Val Gly Met Glu Gln Leu Thr Ser Ser Ser Val Ser Asn Ala Leu
            -55                 -50                 -45

CCA GTC TCA GGA AGT CAC CTG GGA TTG GCT GCC TCA CCC ACT CAC AGT      146
Pro Val Ser Gly Ser His Leu Gly Leu Ala Ala Ser Pro Thr His Ser
        -40                 -35                 -30

GCC ATC CCT GCC CCA GGC CTC CCA GTG GCA ATT CCA AAC CTG GGT CCC      194
Ala Ile Pro Ala Pro Gly Leu Pro Val Ala Ile Pro Asn Leu Gly Pro
    -25                 -20                 -15

TCC CTG AGC TCT CTG CCT TCT GCT CTG TCT TTA ATG CTA CCA ATG GGT      242
Ser Leu Ser Ser Leu Pro Ser Ala Leu Ser Leu Met Leu Pro Met Gly
-10                 -5                      1               5

DTT GGG GAT CGA GGG GTG ATG TGT GGG TTA                              272
Xaa Gly Asp Arg Gly Val Met Cys Gly Leu
            10                  15
```

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 127 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 62..118
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.6
seq IWNLFSLFSTSTT/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

```
ACATCCTTGA TTCTTTACTT TCTCTTAACA CCCTGTATCC AGCTGGTCAT AAATCTAGCA     60

G ATG CTA CAT TCA GAT AAC ATC TGG AAT CTA TTT TCC CTA TTT TCT ACT    109
  Met Leu His Ser Asp Asn Ile Trp Asn Leu Phe Ser Leu Phe Ser Thr
                  -15                 -10                 -5

TCT ACT ACC CTG CCC CGG                                              127
Ser Thr Thr Leu Pro Arg
            1
```

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 135 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 4..75
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.6
seq FHSAAGWSGGGQA/CG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

```
ATT ATG CAA CCC GCC TCC CCG CCC GCC CGG TGG AGC TTC CAC TCG GCT      48
    Met Gln Pro Ala Ser Pro Pro Ala Arg Trp Ser Phe His Ser Ala
                    -20                 -15                 -10

GCG GGC TGG AGC GGC GGC GGG CAG GCG TGC GGA GGA CAC TCC TGC GAC      96
Ala Gly Trp Ser Gly Gly Gly Gln Ala Cys Gly Gly His Ser Cys Asp
                -5                  1                   5

CAG GTA CTG GCT GTG ATC GAA CTT CTC AAC CCT CTC AGG                 135
Gln Val Leu Ala Val Ile Glu Leu Leu Asn Pro Leu Arg
        10                  15                  20
```

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 138..191
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.5
seq LLAGSISHMFSQA/LP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

```
ATATTTTGCC TTAGCAGCCT GGGCTTCGGA ACTGGGCTTG CCCTGTAGCA GGTCACTTGC    60
```

```
ACCTTTCTGC CACAGATGAC GGAAACATTT AAAGTTATGG ATTGTGTCTC TGCATCCTCT  120

TCCCTTCACA CCAGCCA ATG TGT TTT TCA TTT CTC TTG GCT GGC TCA ATT     170
                   Met Cys Phe Ser Phe Leu Leu Ala Gly Ser Ile
                               -15                 -10

TCC CAC ATG TTC TCC CAA GCT CTT CCT CTC CAC TCC CCA GGG CTT CCC   218
Ser His Met Phe Ser Gln Ala Leu Pro Leu His Ser Pro Gly Leu Pro
        -5                  1               5

ACC ACA AAC CGC ACG
Thr Thr Asn Arg Thr                                               233
 10
```

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 214 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Prostate

(ix) FEATURE:
- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 137..199
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.5
  seq SILFHCSVCLFLC/QY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

```
ATATGGCAAG AGATAGAGAT CTAGTTTCAT TCTTCTGCAT ATGGATATCC AATTTTCCCA   60

GCACCATTTA TTGAAGAGAC AGTCCTTTTG CCAGTKTATG TTCTTGGCAA CTTTGTTGAA  120

AATGCATTTA CTGTAG ATG TAT GGA TTC ATT ATT GGG TTA TCT ATT CTG TTC  172
                  Met Tyr Gly Phe Ile Ile Gly Leu Ser Ile Leu Phe
                      -20                 -15                 -10

CAT TGT TCT GTG TGT CTG TTT TTA TGC CAG TAC CAT GCC TGG           214
His Cys Ser Val Cys Leu Phe Leu Cys Gln Tyr His Ala Trp
                -5                  1               5
```

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 231 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 139..210
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.5
seq SLLGCXLAININT/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

```
ATCCTATTGT GTCGTGTAGC TTGTTCTCTA TTTTATAGGT CATTTAAAAT AAAACTCACC   60

TTTGACTTTG TTTAGTCTCT GTTACATGTT TGCTTTTTGT TTCGTTTATG TTTGTACATT  120

TCTCATGTKT TTCTKKCT ATG TCT TTT GGT KGT ATT CTA ACT TTT AGA GTC    171
                    Met Ser Phe Gly Xaa Ile Leu Thr Phe Arg Val
                                    -20                 -15

TCT TTA TTG GGA TGT CNT CTA GCG ATA AAT ATA AAT ACA TTT CCC TCT    219
Ser Leu Leu Gly Cys Xaa Leu Ala Ile Asn Ile Asn Thr Phe Pro Ser
            -10                 -5                   1

AAC AAC CAC TTG                                                    231
Asn Asn His Leu
    5
```

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 269 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 12..77
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.4
seq LGRLCAGSSGVXG/AR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

```
AAAAGCGAGC C ATG GCT GTC TAC GTC GGG ATG CTG CGC CTG GGG AGG CTG    50
             Met Ala Val Tyr Val Gly Met Leu Arg Leu Gly Arg Leu
```

```
                -20                      -15                      -10

TGC GCC GGG AGC TCG GGG GTG STG GGG GCC CGG GCC GSC CTC TCT CGG      98
Cys Ala Gly Ser Ser Gly Val Xaa Gly Ala Arg Ala Xaa Leu Ser Arg
                 -5                      1                   5

AGT TGG CAG GAA GCC AGG TTG CAG GGT GTC CGC TTC CTC AGT TCC AGA     146
Ser Trp Gln Glu Ala Arg Leu Gln Gly Val Arg Phe Leu Ser Ser Arg
         10                  15                  20

GAG GTG GAT CGC ATG GTC TCC ACG CCC ATC GGA GGC CTC AGC TAC GTT     194
Glu Val Asp Arg Met Val Ser Thr Pro Ile Gly Gly Leu Ser Tyr Val
     25                  30                  35

CAG GGG TGC ACC AAA AAG CAT CTT AAC AGC AAG ACT GTG GGC CAG TGC     242
Gln Gly Cys Thr Lys Lys His Leu Asn Ser Lys Thr Val Gly Gln Cys
 40                  45                  50                  55

CTG GAG ACC ACA GCA CAG AGG GTC CCG                                 269
Leu Glu Thr Thr Ala Gln Arg Val Pro
                 60
```

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 276 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 187..255
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 4.4
 seq LVSIFFFWEVTNA/FL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

```
AGATAATTTT GATGAAACCA AGAGGCACGT CTTTCTACAT ACTTCTCTTC ATCKYCMWTT    60

CCTAGTGTTT TWGTTTATKT TTTTTAAATA ATGCCCATGT CTCCTGCTGT CATTCTCTGA   120

GACCACCAAA TAGTTTAATA CCTGGAGTCA GAGATAAGAA TAAACAGGCT TAAGATACTT   180

TAAATA ATG TTC AAT ACT ATA TAC TTG GTC ATA TCA TTA GTG AGC ATA      228
       Met Phe Asn Thr Ile Tyr Leu Val Ile Ser Leu Val Ser Ile
                   -20                     -15                 -10

TTT TTC TTT TGG GAA GTA ACT AAT GCT TTC CTT AAG GCC AGG CGT TGG     276
Phe Phe Phe Trp Glu Val Thr Asn Ala Phe Leu Lys Ala Arg Arg Trp
                 -5                      1                   5
```

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 137 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 36..101
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.4
seq SLPLTTGSSWSLS/SQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

```
ACCTTCTCAA GAACTGTGTT CACCCACTTC CCCAC ATG GCC CTT CCA CCC AAG      53
                                      Met Ala Leu Pro Pro Lys
                                                  -20

GGA TGT GGT AGT CTC CCT TTG ACT ACT GGG TCT TCC TGG AGC CTT TCT    101
Gly Cys Gly Ser Leu Pro Leu Thr Thr Gly Ser Ser Trp Ser Leu Ser
    -15                 -10                     -5

TCT CAA ATA GGA AGC CCT GCT ATT TCC AAC CCT AGG                    137
Ser Gln Ile Gly Ser Pro Ala Ile Ser Asn Pro Arg
  1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 127 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 44..91
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.3
seq FLSWASFLAPLLR/SP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

```
GTCATTTGTC CGTTTCTTCC CCCTTGCCAA TTTTTTAATT AGA ATG TTT GTC TTT      55
                                                Met Phe Val Phe
                                                    -15

TTG TCT TGG GCA AGT TTC TTA GCC CCT CTA CTG AGG AGC CCA TTT CTT      103
Leu Ser Trp Ala Ser Phe Leu Ala Pro Leu Leu Arg Ser Pro Phe Leu
        -10                 -5                  1

CAT TGT CTA ATG GGG ATG CCA GGG                                      127
His Cys Leu Met Gly Met Pro Gly
 5                  10
```

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 302 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 150..233
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.3
seq LLSCSPLXPLGKS/GF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

```
AAKAGTCAGC AGGAGTKAGT TCAGGAATCC TCGGGACAAG GCACTTTCCT GAGCACTGGA    60

CCAGCGACCT CTTGGCTTCC AGGGAGGACA CACAGCCATC ATGGWACCCA THTCTCAGAA    120

GAGTCCAGGC AAACAGTTTA CATTTTCTT ATG AWA ATG AAG TCT GCA AAC AAG      173
                                Met Xaa Met Lys Ser Ala Asn Lys
                                            -25

ATT ACT TTA TTA ART CAC CAC CTT CTC AGC TGT TCT CCT CTG TGW CCT      221
Ile Thr Leu Leu Xaa His His Leu Leu Ser Cys Ser Pro Leu Xaa Pro
-20                 -15                 -10                 -5

CTT GGA AAA AGC GGT TTT TCA TCC TGT CAA AGG CTG GGG AAA AGA GCT      269
Leu Gly Lys Ser Gly Phe Ser Ser Cys Gln Arg Leu Gly Lys Arg Ala
                1                   5                   10

TTA GTC TTT CCT ATT ATR AAG NCC ATC ATC ACC                          302
Leu Val Phe Pro Ile Xaa Lys Xaa Ile Ile Thr
        15                  20
```

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 251 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 150..245
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.2
seq SFLLLFIVIPQTP/RP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

```
AATTTGATAA CATCAGCTAA TATTTTTCAA AGTTAGATTT TTGAGGTATA ATTTACATAA    60

GAGTTACTCT TTCTAGAGGT ATAGTTGAAT GCATTTTCAC AAATGTGTAC AATTGGATAA   120

CCACCAMCAT WAWTCTAGAW ATATAGGTA ATG TGT AAT TAT AAT ATA TAT GTA     173
                                Met Cys Asn Tyr Asn Ile Tyr Val
                                        -30                 -25

CTA TAT AAT ATA GGA TAT TTA TAC CAC CCA AAA AGT TTT CTC TTG CTT     221
Leu Tyr Asn Ile Gly Tyr Leu Tyr His Pro Lys Ser Phe Leu Leu Leu
            -20                 -15                 -10

TTT ATA GTC ATT CCC CAA ACC CCA CGT CCG                             251
Phe Ile Val Ile Pro Gln Thr Pro Arg Pro
             -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 383 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide

(B) LOCATION: 84..164
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.2
seq PLLAAPLLRSLLP/RX

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

```
AACTGAACAG CGGASCGGAC GGGGATCGCC GGCGGGCGGC AAGCGGAGGC GGCCCAGRGC    60

CCGGCGGTCT CCGAGATGTC ACG ATG GCT GTG GCC ATG GTC AAA CTG TGT GAA   113
                          Met Ala Val Ala Met Val Lys Leu Cys Glu
                                  -25                     -20

AGA GCG GGT CTG CCG CTA CTT GCT GCA CCA CTA CTT AGG TCA CTT CTT     161
Arg Ala Gly Leu Pro Leu Leu Ala Ala Pro Leu Leu Arg Ser Leu Leu
        -15                 -10                 -5

CCA AGA GMA CCT CAG CCT GGA CCA GCT CAG CCT CGA TCT GTA CAA GGG     209
Pro Arg Xaa Pro Gln Pro Gly Pro Ala Gln Pro Arg Ser Val Gln Gly
    1               5                   10                  15

CAG CGT TGC CCT GCG AGA CAT CCA CCT GGA AAT CTG GTC TGT GAA CGA     257
Gln Arg Cys Pro Ala Arg His Pro Pro Gly Asn Leu Val Cys Glu Arg
                20                  25                  30

GGT GCT RGA GTC AAT GGA GTC ACC GCT GGA GCT RGT GGA ARG CTT CGT     305
Gly Ala Xaa Val Asn Gly Val Thr Ala Gly Ala Xaa Gly Xaa Leu Arg
            35                  40                  45

GGG CTC CAT CGA GGT DGC CGT GCC CTG GGC TGC TCT GCT CAC CGA CCA     353
Gly Leu His Arg Gly Xaa Arg Ala Leu Gly Cys Ser Ala His Arg Pro
        50                  55                  60

MTG CAC AGT GCG CGT GTC CGG CCT CCA GCT                             383
Xaa His Ser Ala Arg Val Arg Pro Pro Ala
    65                  70
```

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 479 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 99..464
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.2
seq DVLLGLLKDVLLA/RP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

```
TAAACTTCTG AAAGAAAGAG AAGATCTTCC TATATGGAAA GAAAAATACT CCTTTATGGA   60

GAACCTGCTT CAAAATCAAA TCGTGATTGT TTCAGGAG ATG CTA AAT GTG GTA AGA  116
                                         Met Leu Asn Val Val Arg
                                                 -120

GCG CTC AGG KTY CCT CAG TGG TGT GCT GAA TAT TGT CTT TCC ATC CAC    164
Ala Leu Arg Xaa Pro Gln Trp Cys Ala Glu Tyr Cys Leu Ser Ile His
    -115                -110                -105

TAC CAG CAC GGG GGC GTG ATA TGC ACA CAG GTC CAC AAG CAG ACT GTG    212
Tyr Gln His Gly Gly Val Ile Cys Thr Gln Val His Lys Gln Thr Val
-100                -95                 -90                 -85

GTC CAG CTC GCC CTG CGG GTG GCG GAT GAA ATG GAT GTT AAC ATT GGT    260
Val Gln Leu Ala Leu Arg Val Ala Asp Glu Met Asp Val Asn Ile Gly
                -80                 -75                 -70

CAT GAG GTT GGC TAC GTG ATC CCT TTC GAG AAC TGC TGT ACC AAC GAA    308
His Glu Val Gly Tyr Val Ile Pro Phe Glu Asn Cys Cys Thr Asn Glu
            -65                 -60                 -55

ACA ATC CTG AGG TAT TGT ACT GAT GAT ATG CTG CAA AGA GAA ATG ATG    356
Thr Ile Leu Arg Tyr Cys Thr Asp Asp Met Leu Gln Arg Glu Met Met
        -50                 -45                 -40

TCC AAT CCT TTT TTG GGT AGC TAT GGG GTC ATC ATC TTA GAT GAT ATT    404
Ser Asn Pro Phe Leu Gly Ser Tyr Gly Val Ile Ile Leu Asp Asp Ile
    -35                 -30                 -25

CAT GAA AGA AGC ATT GCA ACT GAT GTG TTA CTT GGA CTT CTT AAA GAT    452
His Glu Arg Ser Ile Ala Thr Asp Val Leu Leu Gly Leu Leu Lys Asp
-20                 -15                 -10                  -5

GTT TTA CTA GCA AGA CCA GAA CTG AAG                                479
Val Leu Leu Ala Arg Pro Glu Leu Lys
                  1               5
```

(2) INFORMATION FOR SEQ ID NO: 145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 107..187
(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.2
seq AGLCIGSTSYVHG/DI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

```
ATTGGGAGCA GCAGCATCTA CTTCACAGAC CAGTGTCCAG TTAATTGTGT TTGTGGCAAT   60

CATCCTACAT AAGGCACCAG CTGCTTTTGG ACTGGTTTCC TTCTTG ATG CAT GCT     115
                                                   Met His Ala
                                                           -25

GGC TTA GAG CGG RAW TCG AWT CAG AAA GCA CTT GCT GGT CTT TGC ATT    163
Gly Leu Glu Arg Xaa Ser Xaa Gln Lys Ala Leu Ala Gly Leu Cys Ile
            -20                 -15                 -10

GGC AGC ACC AGT TAT GTC CAT GGT GAC ATA CTT AGG ACT GAG CGG        208
Gly Ser Thr Ser Tyr Val His Gly Asp Ile Leu Arg Thr Glu Arg
             -5                   1               5
```

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 285 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 151..255
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.2
seq LLGSLSLWRWSAM/EP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

```
AATTGCTGGG CTCGAAGCAC AGGAGAGACC AGTCCTTCCT TGTCTCCACT GGGCTGTKTA   60

GTGCTTCTTT CCCAAGGACK TCCATCCCTT CCCCAGGCTT TATGGTTCCA GTKCTTCTAC  120

CATTCTGGAA GCTCCCTAGA ATCTCCTGGA ATG CTT AAT GGA CCT TTC CAG CAC   174
                                 Met Leu Asn Gly Pro Phe Gln His
                                 -35                 -30

CGA AAT TCA AGA ATT ATG ACT CAT CGG TCA GCA GAA AAG ACC CTG CTG    222
Arg Asn Ser Arg Ile Met Thr His Arg Ser Ala Glu Lys Thr Leu Leu
        -25                 -20                 -15

GGA TCT TTG AGC TTG TGG AGG TGG TCG GCA ATG GAA CCT ACG GAC AGG    270
Gly Ser Leu Ser Leu Trp Arg Trp Ser Ala Met Glu Pro Thr Asp Arg
    -10                 -5                   1               5
```

```
TGT ACA AGG GTA GGG                                                285
Cys Thr Arg Val Gly
                10
```

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 409 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 44..175
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4.1
seq IAVGLTCQHVSHA/IS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

```
AAGGTTGTAG ACGCTGCGGC CCGGCCCGGC GGGTAAATAA CAG ATG CGG GTG AAA      55
                                            Met Arg Val Lys

GAT CCA ACT AAA GCT TTA CCT GAG AAA GCC AAA AGA AGT AAA AGG CCT     103
Asp Pro Thr Lys Ala Leu Pro Glu Lys Ala Lys Arg Ser Lys Arg Pro
-40                 -35                 -30                 -25

ACT GTA CCT CAT GAT GAA GAC TCT TCA GAT GAT ATT GCT GTA GGT TTA     151
Thr Val Pro His Asp Glu Asp Ser Ser Asp Asp Ile Ala Val Gly Leu
                -20                 -15                 -10

ACT TGC CAA CAT GTA AGT CAT GCT ATC AGC GTG AAT CAT GTA AAG AGA     199
Thr Cys Gln His Val Ser His Ala Ile Ser Val Asn His Val Lys Arg
            -5                   1               5

GCA ATA GCT GAG AAT CTG TGG TCA GTT TGC TCA GAA TGT TTA AAA GAA     247
Ala Ile Ala Glu Asn Leu Trp Ser Val Cys Ser Glu Cys Leu Lys Glu
    10                  15                  20

AGA AGA TTC TAT GAT GGG CAG CTA GTA CTT ACT TCT GAT ATT TGG TTG     295
Arg Arg Phe Tyr Asp Gly Gln Leu Val Leu Thr Ser Asp Ile Trp Leu
 25                  30                  35                  40

TGC CTC AAG TGT GGC TTC CAG GGA TGT GGT AAA AAC TCA GAA AGC CAA     343
Cys Leu Lys Cys Gly Phe Gln Gly Cys Gly Lys Asn Ser Glu Ser Gln
                45                  50                  55

CAT TCA TTG AAG CAC TTT AAG AGT TCC AGA ACA GAG CCC CAT TGT ATT     391
His Ser Leu Lys His Phe Lys Ser Ser Arg Thr Glu Pro His Cys Ile
            60                  65                  70
```

```
ATA ATT AAT CTG AGC ACA                                          409
Ile Ile Asn Leu Ser Thr
        75
```

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 279 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 184..267
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4
seq FSLLALSMLKGTG/KV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

```
ACATAATCGG CCTTTATGTT ACACTGCCTG GCCAGCCCCT GTTATTCTAG TGCATAATTG    60

ATGGTGCTCA CAAGTGGAAA AGTTAGAAAA GCGGAAGTAA TGTGACGCAG CAGTGCCATG   120

RAGCSSCCGG DVCCCCGGCA GTGAGGGCAA TGCAGAGATG GGCTGCTGCT GGCTACCGCC   180

AGG ATG CCT CAG AAG GGC CTG GGC TTA CTT GGC ATC TTG TCA GGA GAC      228
    Met Pro Gln Lys Gly Leu Gly Leu Leu Gly Ile Leu Ser Gly Asp
                -25                 -20                 -15

TTT TCC CTT CTT GCT TTG TCC ATG CTG AAA GGG ACA GGA AAG GTA GGC      276
Phe Ser Leu Leu Ala Leu Ser Met Leu Lys Gly Thr Gly Lys Val Gly
            -10                 -5                  1

GGG                                                                  279
Gly
```

(2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 326 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 69..233
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4
seq AALCGISLSQLFP/EP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

```
AAGAACCTGA GCAGCCTGTC TTCAGACAGA GAGAGGCCCA CGGCTGTTTC TTGAAAYTGG   60

CGCTGGGA ATG GCC ATG TGG AAC AGG CCA TGB BAG ANG CTG CCT CAG CAG   110
         Met Ala Met Trp Asn Arg Pro Xaa Xaa Xaa Leu Pro Gln Gln
         -55                 -50                 -45

CCT CTS STA GCT GAG CCC ACT GCA GAG GGG GAG CCA CAC CTG CCC ACG   158
Pro Leu Xaa Ala Glu Pro Thr Ala Glu Gly Glu Pro His Leu Pro Thr
    -40                 -35                 -30

GGC CGG GAS BYG ACT GAG GCC AAC CGC TTC GCC TAT GCT GCC CTC TGT   206
Gly Arg Xaa Xaa Thr Glu Ala Asn Arg Phe Ala Tyr Ala Ala Leu Cys
-25                 -20                 -15                 -10

GGC ATC TCC CTG TCC CAG TTA TTT CCT GAA CCC GAA CAC AGC TCC TTC   254
Gly Ile Ser Leu Ser Gln Leu Phe Pro Glu Pro Glu His Ser Ser Phe
                -5                  1               5

TGC ACA GAG TTC ATG GCA GGC CTG GTG SKM TGG CTG GAG TTG TCT GAA   302
Cys Thr Glu Phe Met Ala Gly Leu Val Xaa Trp Leu Glu Leu Ser Glu
        10                  15                  20

GCT GTC TTG CCA ACC ATG ACT GCT                                   326
Ala Val Leu Pro Thr Met Thr Ala
    25                  30
```

(2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 194 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 126..182
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4
seq LLLSPWVTVPVWS/SS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

```
CCTAGTGCTT AAGGGGATTT AGCATCATCC AAGCAGGGTA AACTTTTGTT TTGTTAAAAG   60

AAAAATGTGT TATTCAAGTT GGTGTCCCCA GTTGTAGCTA ACACATCTGG AATGCACTAA  120

CCAAA ATG CTG TGC TTT GGA GAC CTG CTT TTG TCA CCG TGG GTA ACC GTT  170
      Met Leu Cys Phe Gly Asp Leu Leu Leu Ser Pro Trp Val Thr Val
                  -15                 -10                 -5

CCC GTC TGG TCC AGT AGC CCG TGG                                    194
Pro Val Trp Ser Ser Ser Pro Trp
                 1
```

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 27..107
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4
seq LIYFLGLAADTYF/RS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

```
AAGTTAGGTT TAAAGTTTCC TCATTA ATG CAG GAA AAT GCT CAT AAC CTG AGG     53
                             Met Gln Glu Asn Ala His Asn Leu Arg
                                         -25                 -20

CTT TTC AAG TGT TTA TTA ATT TAC TTT CTG GGG CTG GCT GCT GAT ACT     101
Leu Phe Lys Cys Leu Leu Ile Tyr Phe Leu Gly Leu Ala Ala Asp Thr
            -15                 -10                 -5

TAT TTC AGA TCA AAG AGA AAG CCT GTG TCT TTC GTA GTT ACT GTG KKG     149
Tyr Phe Arg Ser Lys Arg Lys Pro Val Ser Phe Val Val Thr Val Xaa
         1               5                  10

CMA GGA AMC TAT GCC ACA GGG                                         170
Xaa Gly Xaa Tyr Ala Thr Gly
 15                 20
```

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 127..303
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 4
seq SVATALFPPLCIS/TG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

```
ACCAAGTCCT CCCAAGTTAT TAACTGGTCA AAAAGGMTTA AAGGMTTAGT TCTTAATAGT   60

TAAGATGCCA CCCATTCAGG GTTTTTTGCT TTCTAAGAGG GAACTTTTAC AGGCATAATT  120

GAGAGA ATG CAT ACA TGC TCT CTA CCT TGT CTT CTC TTT GCT CAG CTG     168
       Met His Thr Cys Ser Leu Pro Cys Leu Leu Phe Ala Gln Leu
                       -55                 -50

CTA GAA TTT TGT AGC TTT CCT CCA GAT GTG CCT CAT AAC TGT GCG CCT     216
Leu Glu Phe Cys Ser Phe Pro Pro Asp Val Pro His Asn Cys Ala Pro
-45                 -40                 -35                 -30

ATT GTC TCA GTC AGG CCG CCT AAT ATT GTA GCA GCC TTT GAA GGG TGC     264
Ile Val Ser Val Arg Pro Pro Asn Ile Val Ala Ala Phe Glu Gly Cys
                -25                 -20                 -15

TCT GTA GCC ACT GCT CTT TTT CCT CCC TTG TGC ATC TCC ACA GGG AAT     312
Ser Val Ala Thr Ala Leu Phe Pro Pro Leu Cys Ile Ser Thr Gly Asn
            -10                  -5                   1

GAG                                                                 315
Glu
```

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
    (A) NAME/KEY: sig_peptide
    (B) LOCATION: 55..138
    (C) IDENTIFICATION METHOD: Von Heijne matrix
    (D) OTHER INFORMATION: score 4
        seq PLLGVLFFQGVYI/VF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

```
AGTCGTTACC GGGAGCTGTA AACAAGGTGT GCAAGCATCT GAAGAGCTGC CGGG ATG      57
                                                            Met

CAG CAG AGA GGA GCA GCT GGA AGC CGT GGC TGC GCT CTC TTC CCT CTG     105
Gln Gln Arg Gly Ala Ala Gly Ser Arg Gly Cys Ala Leu Phe Pro Leu
        -25                 -20                 -15

CTG GGC GTC CTG TTC TTC CAG GGT GTT TAT ATC GTC TTT TCC TTG GAG     153
Leu Gly Val Leu Phe Phe Gln Gly Val Tyr Ile Val Phe Ser Leu Glu
    -10                 -5                   1               5

ATT CGT GCA GAT GCC CAT GTC CGA GGT TAT GTT GGA GAA AAG ATC AAG     201
Ile Arg Ala Asp Ala His Val Arg Gly Tyr Val Gly Glu Lys Ile Lys
                10                  15                  20

TTG AAA TGC ACT TTC AAG TCA ACT TCA GAT GTC ACT GAC AAG CTT ACT     249
Leu Lys Cys Thr Phe Lys Ser Thr Ser Asp Val Thr Asp Lys Leu Thr
            25                  30                  35

ATA GAC TGG ACA TAT CGC CCT CCC AGC AGC AGC CAC ACA GTA TCA ATA     297
Ile Asp Trp Thr Tyr Arg Pro Pro Ser Ser Ser His Thr Val Ser Ile
        40                  45                  50

KTK CAT TAT CAG TCT TTC CAG TAC CCA ACC ACA GCA GGC ACA TTT         342
Xaa His Tyr Gln Ser Phe Gln Tyr Pro Thr Thr Ala Gly Thr Phe
    55                  60                  65
```

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 429 base pairs
    (B) TYPE: NUCLEIC ACID
    (C) STRANDEDNESS: DOUBLE
    (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Homo Sapiens
    (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
    (A) NAME/KEY: sig_peptide
    (B) LOCATION: 109..225
    (C) IDENTIFICATION METHOD: Von Heijne matrix
    (D) OTHER INFORMATION: score 3.9
        seq LILNRSLPTASSS/SS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

```
AAAAATGTAC TGAATGTCCA CTTTGGGCCA GGCTGGGCAC CGAGGACACA GGGGAACTAA   60

GACACAGTCC TGGTCACTGG GAAACTCACA GCCTGTTGGG AAAGAAAG ATG CAM GAV   117
                                                     Met Xaa Xaa

AGT ATC TTC ATT TCA GAA AAA TAT GGA CTC TGC CCG TCT AAG ACT CCC    165
Ser Ile Phe Ile Ser Glu Lys Tyr Gly Leu Cys Pro Ser Lys Thr Pro
    -35                 -30                 -25

ATA ATG AAA ATG CTC CCG TCT TTG ATC CTG AAC CGG TCA TTG CCC ACT    213
Ile Met Lys Met Leu Pro Ser Leu Ile Leu Asn Arg Ser Leu Pro Thr
-20                 -15                 -10                 -5

GCT TCA AGC AGT TCC AGC AGG AAG GAC TTC CGC CTG CCT CAG ACC CGC    261
Ala Ser Ser Ser Ser Ser Arg Lys Asp Phe Arg Leu Pro Gln Thr Arg
                  1               5                  10

CGG CGA ATC ATC ATG GTG CCT CGC AAG GAG GAT CAG ACA CCC CTT AAT    309
Arg Arg Ile Ile Met Val Pro Arg Lys Glu Asp Gln Thr Pro Leu Asn
        15                  20                  25

CCT GCA TCC CAA CCT CAG GCT CCC CCA AAG CCC ATC CCC AGC TTS AAA    357
Pro Ala Ser Gln Pro Gln Ala Pro Pro Lys Pro Ile Pro Ser Xaa Lys
    30                  35                  40

AGT YTG GAA GCT AGM GAT AYC AAS RGC AGC CAG AGG ACA BGA AGA CCT    405
Ser Leu Glu Ala Xaa Asp Xaa Xaa Xaa Ser Gln Arg Thr Xaa Arg Pro
45                  50                  55                  60

GGG CTG AGC AGA GGT CGA AGC TGC                                    429
Gly Leu Ser Arg Gly Arg Ser Cys
                65
```

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 351 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 1..350
(C) IDENTIFICATION METHOD: fasta
(D) OTHER INFORMATION: identity 99.1
region 18..366
id D83597
vrt

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 127..186
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.9
seq FFWVVLFSAGCKV/IT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

```
ATTTCTTGTT CCAAGATCAC CCTTCTGAGT ACCTCTCTGG CTGCCAAATT GCCAGGGCCT    60

TCACAGTTTG ATTCCATTTC TCAGCTCCAA GCATTAGGTA AACCCACCAA GCAATCCTAG   120

CCTGTG ATG GCG TTT GAC GTC AGC TGC TTC TTT TGG GTG GTG CTG TTT      168
       Met Ala Phe Asp Val Ser Cys Phe Phe Trp Val Val Leu Phe
       -20                 -15                 -10

TCT GCC GGC TGT AAA GTC ATC ACC TCC TGG GAT CAG ATG TAC ATT GAG     216
Ser Ala Gly Cys Lys Val Ile Thr Ser Trp Asp Gln Met Tyr Ile Glu
     -5                   1               5                  10

AAA GAA GCC AAC AAA ACA TAT AAC TGT GAA AAT TTA GGT CTC AGT GAA     264
Lys Glu Ala Asn Lys Thr Tyr Asn Cys Glu Asn Leu Gly Leu Ser Glu
                15                  20                  25

ATC CCT GAC ACT CTA CCA AAC ACA ACA GAA TTT TTG GAA TTC AGC TTT     312
Ile Pro Asp Thr Leu Pro Asn Thr Thr Glu Phe Leu Glu Phe Ser Phe
            30                  35                  40

AAT TTT TTG CCT ACA ATT CAC AAT AGA ACC TCC AGC AGG                 351
Asn Phe Leu Pro Thr Ile His Asn Arg Thr Ser Ser Arg
        45                  50                  55
```

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 410 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 96..383
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.9
seq IMNLTVMLDTAXG/KX

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

```
AGAACCACGC CGCTTCCCAG TCTCTGTGCG AGGCGTGAAG CGCGGACCTT TCAACAAGGG    60
```

```
CTTTATTAAT TCTCACGCTG CGGCCCTGGA AAGCG ATG GAG GTG GCG GCT AAT      113
                                       Met Glu Val Ala Ala Asn
                                               -95

TGC TCC CTA CGG GTG AAG AGA CCT CTG TTG GAT CCC CGC TTC GAG GGT     161
Cys Ser Leu Arg Val Lys Arg Pro Leu Leu Asp Pro Arg Phe Glu Gly
-90                 -85                 -80                 -75

TAC AAG BTC TCT CTT GAG CCG CTG CCT TGT TAC CAG CTG GAG CTT GAC     209
Tyr Lys Xaa Ser Leu Glu Pro Leu Pro Cys Tyr Gln Leu Glu Leu Asp
                -70                 -65                 -60

GCA GCT GTG GCA KAG GTA AAA CTT CGA GAT GAT CAA TAT ACA CTG GAA     257
Ala Ala Val Ala Xaa Val Lys Leu Arg Asp Asp Gln Tyr Thr Leu Glu
            -55                 -50                 -45

CAC ATG CAT GCT TTT GGA ATG TAT AAT TAC CTG CAC TGT GAT TCA TGG     305
His Met His Ala Phe Gly Met Tyr Asn Tyr Leu His Cys Asp Ser Trp
        -40                 -35                 -30

TAT CAA GAC AGT GTC TAC TAT ATT GAT ACC CTT GGA AGA ATT ATG AAT     353
Tyr Gln Asp Ser Val Tyr Tyr Ile Asp Thr Leu Gly Arg Ile Met Asn
    -25                 -20                 -15

TTA ACA GTA ATG CTG GAC ACT GCC TTW GGR AAA MCA CGA GAG GTG TTT     401
Leu Thr Val Met Leu Asp Thr Ala Xaa Gly Lys Xaa Arg Glu Val Phe
-10                 -5                   1               5

CGA CTC CTA                                                         410
Arg Leu Leu
```

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 347 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 63..179
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.9
  seq VLAIGLLHIVLLS/IP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

```
AGGGAACCGA TCCCGGGCCG TTGATCTTCG GCCCCACACG AACAGCAGAG AGGGGCATCA    60

GG ATG AAT GTK GGC ACA GCG CAC AGS DAG GTG AAC CCC AAC ACG CGG      107
```

```
    Met Asn Val Gly Thr Ala His Xaa Xaa Val Asn Pro Asn Thr Arg
                    -35                 -30                 -25

GTK ATG AAC AGC CGT GGC ATC TGG CTC TCC TAC GTG CTG GCC ATC GGT      155
Val Met Asn Ser Arg Gly Ile Trp Leu Ser Tyr Val Leu Ala Ile Gly
                -20                 -15                 -10

CTC CTC CAC ATC GTG CTG CTG AGC ATC CCG TTT GTK AGT GTC CCT GTC      203
Leu Leu His Ile Val Leu Leu Ser Ile Pro Phe Val Ser Val Pro Val
             -5                   1               5

GTC TGG ACC CTC ACC AAC CTC ATT CAC AAC ATG GGC ATG TAT ATC TTC      251
Val Trp Thr Leu Thr Asn Leu Ile His Asn Met Gly Met Tyr Ile Phe
     10                  15                  20

CTG CAC ACG GTG AAG GGG WCA CCC TTT GAG ACC CCG GAC CAG GGC AAG      299
Leu His Thr Val Lys Gly Xaa Pro Phe Glu Thr Pro Asp Gln Gly Lys
 25                  30                  35                  40

GCG AGG CTG CTW WCC CAC TGK TDA GCA GAT GGA TTA TGG GGT CCA GTT      347
Ala Arg Leu Leu Xaa His Xaa Xaa Ala Asp Gly Leu Trp Gly Pro Val
                 45                  50                  55
```

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 151 base pairs
  (B) TYPE: NUCLEIC ACID
  (C) STRANDEDNESS: DOUBLE
  (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
  (A) ORGANISM: Homo Sapiens
  (F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
  (A) NAME/KEY: sig_peptide
  (B) LOCATION: 8..76
  (C) IDENTIFICATION METHOD: Von Heijne matrix
  (D) OTHER INFORMATION: score 3.9
      seq SWWTLLSSSPSFM/IS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

```
ATTTATT ATG GAA AAC TTT AAC ATG TAT AAA AAT AAG AGC TGG TGG ACC      49
        Met Glu Asn Phe Asn Met Tyr Lys Asn Lys Ser Trp Trp Thr
                    -20                 -15                 -10

CTT TTG TCC TCA TCA CCC AGC TTT ATG ATC AGT TTT GTT TCA TCT GTA      97
Leu Leu Ser Ser Ser Pro Ser Phe Met Ile Ser Phe Val Ser Ser Val
                 -5                   1               5

CTA CCA GTG CTA CTT ACC ATC TCT AGG TTC ATT TTG AAG CAA ATC CCA      145
Leu Pro Val Leu Leu Thr Ile Ser Arg Phe Ile Leu Lys Gln Ile Pro
         10                  15                  20
```

```
GAC CAG                                                             151
Asp Gln
     25
```

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 351 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 142..258
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.9
seq VLAIGLLHIVLLS/IP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

```
AGATTCGGCC GGAGCTGCCA GCGGGGAGGC TGCAGCCGCG GGTTGTTACA GCTGCTGGAG   60

CAGCAGCGGC CCCCGCTCCC GGGAACCGKT CCCGGGCCGT TGRTCTTCGG CCCCACACGA  120

ACAGCAGAGA GGGGCAGCAG G ATG AAT GTG GGS ACA GND CAC AGC GAG GTG     171
                        Met Asn Val Gly Thr Xaa His Ser Glu Val
                                        -35                 -30

AAC CCC AAC ACG CGG GTG ATG AAC AGC CGT GGG ATC TGG CTC TCC TAC     219
Asn Pro Asn Thr Arg Val Met Asn Ser Arg Gly Ile Trp Leu Ser Tyr
            -25                 -20                 -15

GTG CTG GCC ATC GGT CTC CTC CAC ATC GTG CTC CTG AGC ATC CCG TTT     267
Val Leu Ala Ile Gly Leu Leu His Ile Val Leu Leu Ser Ile Pro Phe
        -10                 -5                   1

GTG AGT GTC CCT GTC GTC TGG ACC CTC ACC AAC CTC ATT CAC AAC ATG     315
Val Ser Val Pro Val Val Trp Thr Leu Thr Asn Leu Ile His Asn Met
     5                  10                  15

GGC ATG TAT ATC TTC CTG TAC ACG GTG AAG GGG ACA                     351
Gly Met Tyr Ile Phe Leu Tyr Thr Val Lys Gly Thr
 20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 234 base pairs
(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 88..129
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.8
seq AAASAVSVLLVAA/ER

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

```
AABGCTTCGT AGTGGAGGAA CGGGTTTGGC GTGTGGGACG CAGCTGCCTC TGTACTGGGG    60

AGTCACGGAG TCCCGGGCTC CAGGGAC ATG GCG GCG GCC TCT GCG GTG TCG GTG   114
                              Met Ala Ala Ala Ser Ala Val Ser Val
                                                  -10

CTG CTG GTG GCG GCG GAG AGG AAC CGG TGG CAT CGT CTC CCG AGC CTG     162
Leu Leu Val Ala Ala Glu Arg Asn Arg Trp His Arg Leu Pro Ser Leu
 -5                  1               5                  10

CTC CTG CCG CCG AGG ACA TGG GTG TGG AGG CAA AGA ACC ATG AAG TAC     210
Leu Leu Pro Pro Arg Thr Trp Val Trp Arg Gln Arg Thr Met Lys Tyr
             15                  20                  25

ACA ACA GCC ACA GGA AGA AAC ATG                                     234
Thr Thr Ala Thr Gly Arg Asn Met
         30                  35
```

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 461 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 177..308
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.8
seq SGSGLSWARLSQS/RS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

```
ACTCTTTGCC ACCCTCAGAG GCGAGCTGTG GAAGCCTTGA CTCTTAGGGC CGTTTTAGAA   60

CCGGGGCCTC GGACCGGCGG GGTTTCTGCA CGTGGAACCG GAACATCTGA GATGATCGSM  120

RGGCCCTGTG GAGTGTGGGG AGCGCGGGAG TTCTTTCTTC CCTCGAGGCC CGTGCC ATG  179
                                                              Met

GCT TAC TCC AAA GCC AGT GGG TCC CCA GTC CTA AGC CAA GCA GTC CCG    227
Ala Tyr Ser Lys Ala Ser Gly Ser Pro Val Leu Ser Gln Ala Val Pro
            -40                 -35                 -30

GGG GAA AAC GCT TCT CAT CGC CGT GGG AGC GCG GAT CTT GGA AGT GGC    275
Gly Glu Asn Ala Ser His Arg Arg Gly Ser Ala Asp Leu Gly Ser Gly
        -25                 -20                 -15

TCT GGG CTT TCT TGG GCG AGG CTC TCA CAG AGT AGA TCG GAA ATC CAT    323
Ser Gly Leu Ser Trp Ala Arg Leu Ser Gln Ser Arg Ser Glu Ile His
    -10                  -5                   1               5

TCT GCT GGC CCG CCC CAC CTC GGA GGA CGG ACT AAC GGA CCT GAG TTC    371
Ser Ala Gly Pro Pro His Leu Gly Gly Arg Thr Asn Gly Pro Glu Phe
                 10                  15                  20

CCG GCC CTA TCT TAC TCT TCT CAG CTT CTC AGC TTG GCT CAG CTC AGA    419
Pro Ala Leu Ser Tyr Ser Ser Gln Leu Leu Ser Leu Ala Gln Leu Arg
            25                  30                  35

GGA AGA GGA ATC ACT GAA GTC TCA GAG AAG TCT CCA CTC ATC            461
Gly Arg Gly Ile Thr Glu Val Ser Glu Lys Ser Pro Leu Ile
        40                  45                  50
```

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 459 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 175..285
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8
  seq RPVLLHLHQTAHA/DE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

```
AGGCTCCGCG CTCTGGAGGC TCAGGCGCCG CGTGGGGCCC GCACCTCTGG GCAGCAGCGG   60
```

```
CAGCCGAGAC TCACGGTCAA GCTAAGGCGA AGAGTGGGTG GCTGAAGCCA TACTATTTTA  120

TAGAATTAAT GGRAARCMHG AAAAGMCATC ACAAACCAAG AAGAACTTTG GAAA ATG    177
                                                            Met

AAG CCT AGG AGA AAT TTA GAA GAA GAC GAT TAT TTG CAT AAG GAC ACG    225
Lys Pro Arg Arg Asn Leu Glu Glu Asp Asp Tyr Leu His Lys Asp Thr
    -35                 -30                     -25

GGA GAG ACC AGC ATG CTA AAA AGA CCT GTG CTT TTG CAT TTG CAC CAA    273
Gly Glu Thr Ser Met Leu Lys Arg Pro Val Leu Leu His Leu His Gln
-20                 -15                     -10                 -5

ACA GCC CAT GCT GAT GAA TTT GAC TGC CCT TCA GAA CTT CAG CAC ACA    321
Thr Ala His Ala Asp Glu Phe Asp Cys Pro Ser Glu Leu Gln His Thr
                1               5                   10

CAG GAA CTC TTT CCA CAG TGG CAC TTG CCA ATT AAA ATA GCT GCT ATT    369
Gln Glu Leu Phe Pro Gln Trp His Leu Pro Ile Lys Ile Ala Ala Ile
        15                  20                  25

ATA GCA TCT CTG ACT TTT CTT TAC ACT CTT CTG AGG GAA GTA ATT CAC    417
Ile Ala Ser Leu Thr Phe Leu Tyr Thr Leu Leu Arg Glu Val Ile His
    30                  35                  40

CCT TTA GCA ACT TCC CAT CAA CAA TAT TTT TAT AAA ATT CAA            459
Pro Leu Ala Thr Ser His Gln Gln Tyr Phe Tyr Lys Ile Gln
45                  50                  55
```

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 25..81
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.7
seq IPCAHMLVCPTIG/DI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

```
AATTTGTAAG AATATTATAT ATAG ATG ATC ATC TGT TAT GAT ATT CCT TGT      51
                           Met Ile Ile Cys Tyr Asp Ile Pro Cys
                                           -15

GCA CAT ATG TTG GTT TGT CCT ACT ATT GGT GAT ATT AAG TTT GAT CAC     99
Ala His Met Leu Val Cys Pro Thr Ile Gly Asp Ile Lys Phe Asp His
```

```
-10               -5                1               5

TTG ATG AAG TGG TAT CCA TCA GAT TTC TCT ACT GAA AGG CTG            141
Leu Met Lys Trp Tyr Pro Ser Asp Phe Ser Thr Glu Arg Leu
            10                  15                  20
```

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 393 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 184..240
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.7
seq STLASVPPAATFG/AD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

```
AACCAGGCTC TATTTAGAGC CGGGTAGGGG AGCGCAGGNC CAGATACCTC AGCGCTACCT    60

GGCGGAACTG GATTTCTCTC CCGCCTGCCG GCCTGCCTGC CACAGCCGGA CTCCGCCACT   120

CCGGTAGCCC CATGGCTGGM AACCTGTGAG ATTAGCAATA TTTTTAGCAA CTACTTCAGT   180

GCG ATG TAC AGC TCG GAG GAC TCC ACC CTG GCC TCT GTT CCC CCT GCT     228
    Met Tyr Ser Ser Glu Asp Ser Thr Leu Ala Ser Val Pro Pro Ala
                -15                 -10                 -5

GCC ACC TTT GGG GCC GAT GAC TTG GTA CTG ACC CTG AGC AAC CCC CAG     276
Ala Thr Phe Gly Ala Asp Asp Leu Val Leu Thr Leu Ser Asn Pro Gln
                1               5                   10

ATG TCA TTG GAG GGT ACA GAG AAG GCC AGC TGG TTG GGG GAA CAG CCC     324
Met Ser Leu Glu Gly Thr Glu Lys Ala Ser Trp Leu Gly Glu Gln Pro
        15                  20                  25

CAG TTC TGG TCG AAG ACG CAG GTT CTG GAC TGG ATC AGC TAC CAA GTG     372
Gln Phe Trp Ser Lys Thr Gln Val Leu Asp Trp Ile Ser Tyr Gln Val
    30                  35                  40

GAG AAG AAC AAG TAC GAC GCG                                         393
Glu Lys Asn Lys Tyr Asp Ala
45                  50
```

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 54..248
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.7
seq QLEGLNWLRFSWA/QG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

```
ACCCTGAATA CGAAGAACAT AAGCAAAGCT ACTGGAGACA CCGAGAACTA ATT ATG      56
                                                           Met
                                                           -65

GGG GAA GAC CCT KCC CAG CCC CGC AAG TAT AAG AAG WWG AAG AWG GAG    104
Gly Glu Asp Pro Xaa Gln Pro Arg Lys Tyr Lys Lys Xaa Lys Xaa Glu
            -60                     -55                     -50

CTA CAG GGT GAT KGG CCT CCC AGT TCT CCC ACT AAT GAT CCT ACC GTG    152
Leu Gln Gly Asp Xaa Pro Pro Ser Ser Pro Thr Asn Asp Pro Thr Val
        -45                     -40                     -35

AAA TAT GAG ACT CAG CCA CGG TTT ATC ACA GCC ACT GGA GGC ACC CTG    200
Lys Tyr Glu Thr Gln Pro Arg Phe Ile Thr Ala Thr Gly Gly Thr Leu
    -30                     -25                     -20

CAC ATG TAT CAG TTG GAA GGG CTG AAC TGG CTA CGC TTC TCC TGG GCC    248
His Met Tyr Gln Leu Glu Gly Leu Asn Trp Leu Arg Phe Ser Trp Ala
    -15                     -10                      -5

CAG GGC ACT KWC GGG                                                263
Gln Gly Thr Xaa Gly
  1               5
```

(2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 148..273
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.7
seq LLGCLQCCWLQSG/RA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

```
ACCAATTTTG TAGTTATCTG ATCTGAAGGA AGATGTGTGT GGAGGTGTTT AGTGATGTTT   60

TCCGATGACG GTGATTCCCC CTAAATCTAC GTATTAAATA CAATGGAACA GGATCCACAG  120

TTCACCCCTA ATAATATAGT TTACTGA ATG TTT TAT GTA GCT ATG ACC AAA ACT  174
                              Met Phe Tyr Val Ala Met Thr Lys Thr
                                      -40                 -35

CAC AAA AGG ATC AGA AGC CTC TGT AAC ATC CAC CAT GGT TTG TTC CAG    222
His Lys Arg Ile Arg Ser Leu Cys Asn Ile His His Gly Leu Phe Gln
            -30                 -25                 -20

TTT ACT CAG CAG CTC CTG GGC TGT CTT CAG TGC TGT TGG CTG CAA TCA    270
Phe Thr Gln Gln Leu Leu Gly Cys Leu Gln Cys Cys Trp Leu Gln Ser
        -15                 -10                  -5

GGC AGA GCC CCA GCT ACC TAT TAC CTT GTG GAG AGT ATT GAA AAG TCA    318
Gly Arg Ala Pro Ala Thr Tyr Tyr Leu Val Glu Ser Ile Glu Lys Ser
      1               5                  10                  15

GCA CAT GGC TCT GTA TTA NGT ACT TAT GAT CAA ACT CAG ACT CGC ATA    366
Ala His Gly Ser Val Leu Xaa Thr Tyr Asp Gln Thr Gln Thr Arg Ile
                 20                  25                  30

GGC AGG                                                            372
Gly Arg
```

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 343 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 158..337
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.7
seq XTCASXNPSQCLA/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

```
ACAGAATCTT TAGGTGGGCC TGTTGGTGAG GTCACTTTTC CCTAATGGTA TATTCCAGTT    60

CCTGTAGATC CTATTCCAGT TCCCAGGACA TATTCCAACC TCGACCTCCA GCCAACTTTG   120

AACCCCTGAA GTTGTGTGCT GATGTGTTTC TAACAAC ATG GTC TCA CCC AAA GAT    175
                                        Met Val Ser Pro Lys Asp
                                        -60                 -55

CTT CCT CTT GTG CTT TTG CAG GAC ATT AAA GTT CCC AGC TCC ATG ACT      223
Leu Pro Leu Val Leu Leu Gln Asp Ile Lys Val Pro Ser Ser Met Thr
                -50                 -45                 -40

GGA TCA CAT GCT GGA AAC CCT CAT ATA GAA AGG AAT GAT CTC CCC AGA      271
Gly Ser His Ala Gly Asn Pro His Ile Glu Arg Asn Asp Leu Pro Arg
            -35                 -30                 -25

CAT GGT TCT CCT CAA TTT TTT ACA GGH HYG ACT TGT GCT TCT RCA AAC      319
His Gly Ser Pro Gln Phe Phe Thr Gly Xaa Thr Cys Ala Ser Xaa Asn
        -20                 -15                 -10

CCA TCT CAG TGT CTG GCA GCA TTT                                      343
Pro Ser Gln Cys Leu Ala Ala Phe
    -5                      1
```

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..45
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.6
seq FXSLFCLYFSCFL/HI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

```
ATG GAA TTT KTT TCT CTT TTC TGT CTC TAC TTC AGC TGT TTC CTA CAT      48
Met Glu Phe Xaa Ser Leu Phe Cys Leu Tyr Phe Ser Cys Phe Leu His
-15                 -10                 -5                  1

ATT ATA TAT TTT KKC AGC TGT TTC CTA TAC                              78
Ile Ile Tyr Phe Xaa Ser Cys Phe Leu Tyr
            5                   10
```

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 10..144
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.6
seq ALLELIDSPECLS/KC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

```
ACTGGGAAG ATG GCG CTG CAC TTC CAG AGT TTG GCT GAA TTG GAA RTG TTA    51
          Met Ala Leu His Phe Gln Ser Leu Ala Glu Leu Glu Xaa Leu
          -45                 -40                 -35

TGT ACT CAT CTC TAC ATA GGG ACT GAT CTT ACA CAA AGA ATA GAG GCT     99
Cys Thr His Leu Tyr Ile Gly Thr Asp Leu Thr Gln Arg Ile Glu Ala
    -30                 -25                 -20

GAG AAA GCA CTC TTG GAA CTT ATT GAC AGT CCA GAA TGT CTC AGC AAG    147
Glu Lys Ala Leu Leu Glu Leu Ile Asp Ser Pro Glu Cys Leu Ser Lys
-15                 -10                 -5                   1

TGT CAA CTT TTA TTA GAA CAA GGA ACA ACA TCC TAT GCT CAG CTC CTT    195
Cys Gln Leu Leu Leu Glu Gln Gly Thr Thr Ser Tyr Ala Gln Leu Leu
                5                   10                  15

GCA GCA ACC GKV                                                    207
Ala Ala Thr Xaa
        20
```

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 418 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 299..379
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.6
seq LTLLLITPSPSPL/LF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

```
ACCTTGGGCT CCAAATTCTA GCTCATAAAG ATGCAAGTKT TGCAATTTCC TATAAATGGT    60

TAAGAAAAGA GCAAGCTGTC CAGAGAGTGA GAAGTTTGAA AAGAGAGGTG CATAAGAGAG   120

AAATGATGTC CATTTGAGCC CCACCACGGA GGTTATGTGG TCCCAAAAGG AATGATGGCC   180

AAGCAATTAA TTTTTCCTCC TAGTTCTTAG CTTGCTTCTG CATTGATTGG CTTTACACAA   240

CTGGCATTTA GTCTGCATTA CACAAATAGA CACTAATTTA TTTGGAACAA GCAGCAAA     298

ATG AGA ACT TTA TTT GGT GCA GTC AGG GCT CCA TTT AGT TCC CTC ACT     346
Met Arg Thr Leu Phe Gly Ala Val Arg Ala Pro Phe Ser Ser Leu Thr
        -25                 -20                 -15

CTG CTT CTA ATC ACC CCT TCT CCC AGC CCT CTT CTA TTT GAT AGA GGT     394
Leu Leu Leu Ile Thr Pro Ser Pro Ser Pro Leu Leu Phe Asp Arg Gly
    -10                 -5                   1               5

CTG TCC CTC AGA TCA GCA ATG TCG                                     418
Leu Ser Leu Arg Ser Ala Met Ser
                10
```

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 238 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 107..229
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.6
seq AVSSLIAVGTSHG/LA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

```
ACACTCCTAT GATACTTCAT CTGTGGCAAG CTCAGATAGT GGTGACAGGA CCAACTTAAA    60
```

```
AAGGAAGAAG AAATTACCTG ATTCTTTTTC ACTTCATGGA TCAGTT ATG CGC CAT      115
                                                  Met Arg His
                                                      -40

TCA CTT TTG AAG GGA ATT TCT GCC CAG ATA GTG TCT GCA GCT GAC AAA     163
Ser Leu Leu Lys Gly Ile Ser Ala Gln Ile Val Ser Ala Ala Asp Lys
            -35                 -30                 -25

GTA GAT GCT GGC TTG CCT ACA GCA ATT GCA GTA TCC AGT CTG ATA GCA     211
Val Asp Ala Gly Leu Pro Thr Ala Ile Ala Val Ser Ser Leu Ile Ala
        -20                 -15                 -10

GTG GGT ACA TCT CAT GGA TTG GCT GGG                                 238
Val Gly Thr Ser His Gly Leu Ala Gly
     -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 188 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 120..164
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.5
seq LSCFIFFYISSLC/CF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

```
TGTGAAGATG ACAGAGATCT AACTTCTGAG AGCAGAGGTG TCAAGTGACG GTCCCCTTGG    60

AGGAATGGTC TTTGCATCTG ACTACTTCCT TCTGCAACTG TGTTCTTCCA TTAGCTTCC    119

ATG ACA CTC TCC TGC TTT ATT TTT TTC TAC ATC TCT AGC CTT TGC TGT     167
Met Thr Leu Ser Cys Phe Ile Phe Phe Tyr Ile Ser Ser Leu Cys Cys
-15                 -10                  -5                   1

TTC CTC TCC TAC CCC ACC AGG                                         188
Phe Leu Ser Tyr Pro Thr Arg
              5
```

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 base pairs

(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 28..72
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.5
seq LCFLLPHHRLQEA/RQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

```
ATAGATCAGT GACGTCTTTT TCTTCAG ATG ATC CTA TGT TTC CTT CTT CCT CAT    54
                              Met Ile Leu Cys Phe Leu Leu Pro His
                              -15                 -10

CAT CGT CTT CAG GAA GCC AGA CAG ATT CAA GTA TTG AAG ATG CTG CCA     102
His Arg Leu Gln Glu Ala Arg Gln Ile Gln Val Leu Lys Met Leu Pro
     -5                  1               5                  10

AGG GAA AAA TTA AGR AGA AGR AGA AGA GAG AAA ACA AAT AAA TGG GAA     150
Arg Glu Lys Leu Arg Arg Arg Arg Arg Glu Lys Thr Asn Lys Trp Glu
                15                  20                  25

AAA AGA AAG GGC AGC GGG                                             168
Lys Arg Lys Gly Ser Gly
            30
```

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 64..105
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.5
seq FSLFALNMPLGFC/VY

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

```
TTTATTTTAA CCATCTTTTA CTATTTTTAG AAGGAAACTA GCTTTAGTAG TGGGTTGCCC    60

TGT ATG TTT TCT CTT TTT GCT CTT AAT ATG CCA TTG GGT TTT TGT GTG     108
    Met Phe Ser Leu Phe Ala Leu Asn Met Pro Leu Gly Phe Cys Val
                    -10                     -5                  1

TAT GTG ATT TTC AAA ATT CAT GAC TGG                                 135
Tyr Val Ile Phe Lys Ile His Asp Trp
              5                  10
```

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 303 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 163..255
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5
  seq SVWGVLPPPACSA/DL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

```
ATTTGATTTT AGTCAGGGTG TAAGAATATG TATTATTGTT CCCAAAAAAA TCTGTGTAAA    60

AACTTCATAG TGTGAAACAG TGGCAACTGS KTGATTAAAA CATCATTTAG AAAAGACACT   120

CTTCCCTGTT TTGAAATTGA CTCCTCAAAA GGACAGCTGA AC ATG GCC TCT TCT      174
                                               Met Ala Ser Ser
                                                   -30

CCA GGT GTC GCC ATG CAC TCC CTC TGG GCC ACC ATA CAC ACT TCT GTG     222
Pro Gly Val Ala Met His Ser Leu Trp Ala Thr Ile His Thr Ser Val
        -25                 -20                 -15

TGG GGC GTG CTC CCA CCT CCA GCC TGC TCA GCT GAT CTT TTG TTC AGC     270
Trp Gly Val Leu Pro Pro Pro Ala Cys Ser Ala Asp Leu Leu Phe Ser
    -10                     -5                   1               5

AAT GCC TGT CTA CTT CCC CAT GAG ATC CAC CTG                         303
Asn Ala Cys Leu Leu Pro His Glu Ile His Leu
                10                  15
```

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 60..194
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.5
seq LPRLLSLSQHSES/WI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

```
AGAGTTTCCG GTCTGGGCTT TGGCGGGTCT GGTTTGAAGC TCTCCTGTTT GACGAAAGT     59

ATG TCT CAG GAA GGT GCG GTC CCA GCT AGC GCG GTT CCC CTG GAA GAA     107
Met Ser Gln Glu Gly Ala Val Pro Ala Ser Ala Val Pro Leu Glu Glu
-45                 -40                 -35                 -30

TTA AGT AGC TGG CCA GAG GAG CTA TGC CGC CGG GAA CTG CCG TCC GTC     155
Leu Ser Ser Trp Pro Glu Glu Leu Cys Arg Arg Glu Leu Pro Ser Val
                -25                 -20                 -15

CTG CCC CGA CTC CTC TCA TTG TCT CAA CAT TCT GAA AGT TGG ATT GAG     203
Leu Pro Arg Leu Leu Ser Leu Ser Gln His Ser Glu Ser Trp Ile Glu
            -10                 -5                   1

CAT ATT CAA ATT TTG AAA ATT ATT GTA GAA ATG TTT TTA CCT CAT ATG     251
His Ile Gln Ile Leu Lys Ile Ile Val Glu Met Phe Leu Pro His Met
    5                   10                  15

AAC CAC CTG ACA TTG GAA CAG ACT TTC TTT TCA CAA GTG TTA CCA AAG     299
Asn His Leu Thr Leu Glu Gln Thr Phe Phe Ser Gln Val Leu Pro Lys
20                  25                  30                  35

ACT GTG AAA TTA TTC GAT                                             317
Thr Val Lys Leu Phe Asp
                40
```

(2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 370 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 254..361
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 3.5
seq AAVVFAVVLSIHA/TV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

```
AGTAACTGTG AGGAAGGCTG CAGAGTGGCG ACGTCTACGC CGTAGGTTGG AGGCTGTGGG   60

GGGTGGCCGG GCGCCAGCTC CCAGGCCGCA GAAGTGACCT GCGGTGGAGT TCCCTCCTCG  120

CTGCTGGAGA ACGGAAGGGA ARAAGGTTSC TGGCCGGGTG AAAGTGCCTC CCTCTGCTTG  180

ACGGGGCTGA GGGGCCCGAA GTCTAGGGCG TCCGTAGTCG CCCCGGCCTC CGTGAAGCCC  240

CAGGTCTAGA GAT ATG ACC CGA GAG TGC CCA TCT CCG GCC CCG GGG CCT     289
               Met Thr Arg Glu Cys Pro Ser Pro Ala Pro Gly Pro
                   -35                 -30                 -25

GGG GCT CCG CTG AGT GGA TCG GTG CTG GCA GAG GCG GCA GTA GTG TTT    337
Gly Ala Pro Leu Ser Gly Ser Val Leu Ala Glu Ala Ala Val Val Phe
                -20                 -15                 -10

GCA GTG GTG CTG AGC ATC CAC GCA ACC GTA TGG                        370
Ala Val Val Leu Ser Ile His Ala Thr Val Trp
             -5                  1
```

(2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 369..470
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 2..103
id AA059664
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 216..269

(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 14.8
seq LLWWALLLGLAQA/CP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

```
AAGTGGATGG TTCCAGGCAC CCTGTCTGGG GCAGGGAGGG CACAGGCCTG CACATCGAAG    60

GTGGGGTGGG ACCAGGCTGC CCCTCGCCCC AGCATCCAAG TCCTCCCTTG GGCGCCCGTG   120

GCCCTGGCAG ACTCTCAGGG CTAAGGTCCT CTGTTGCTTT TTGGTTCCAC CTTAGAAGAG   180

GCTCGCTTGA CTAAGAGTAG CTTGAAGGAG GCACC ATG CAG GAG CTG CAT CTG      233
                                      Met Gln Glu Leu His Leu
                                                      -15

CTC TGG TGG GCG CTT CTC CTG GGC CTG GCT CAG GCC TGC CCT GAG CCC     281
Leu Trp Trp Ala Leu Leu Leu Gly Leu Ala Gln Ala Cys Pro Glu Pro
        -10                 -5                  1

TGC GAC TGT GGG GAA AAG TAT GGC TTC CAG ATC GCC GAC TGT GCC TAC     329
Cys Asp Cys Gly Glu Lys Tyr Gly Phe Gln Ile Ala Asp Cys Ala Tyr
  5                  10                  15                  20

CGC GAC CTA GAA TCC GTG CCG CCT GGC TTC CCG GCC AAT GTG ACT ACA     377
Arg Asp Leu Glu Ser Val Pro Pro Gly Phe Pro Ala Asn Val Thr Thr
                 25                  30                  35

CTG AGC CTG TCA GCC AAC CGG CTG CCA GGC TTG CCG GAR GGT GCC TTC     425
Leu Ser Leu Ser Ala Asn Arg Leu Pro Gly Leu Pro Glu Gly Ala Phe
             40                  45                  50

AGG GAG GTG CCC CTG CTG CAG TCG CTG TGG CTG GCA CAC AAT GAG         470
Arg Glu Val Pro Leu Leu Gln Ser Leu Trp Leu Ala His Asn Glu
         55                  60                  65
```

(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 331 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 69..328
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 95
region 1..260
id H96534
est

```
    (ix) FEATURE:
          (A) NAME/KEY: sig_peptide
          (B) LOCATION: 14..67
          (C) IDENTIFICATION METHOD: Von Heijne matrix
          (D) OTHER INFORMATION:  score 13.6
                                  seq LLLLALCATGAQG/LY

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

CTCTCTGCGG GCG ATG GGG CGG CAG GCC CTG CTG CTT CTC GCG CTG TGC        49
               Met Gly Arg Gln Ala Leu Leu Leu Leu Ala Leu Cys
                           -15                 -10

GCC ACA GGC GCC CAG GGG CTC TAC TTC CAC ATC GGC GAG ACC GAG AAG       97
Ala Thr Gly Ala Gln Gly Leu Tyr Phe His Ile Gly Glu Thr Glu Lys
     -5                  1               5                  10

CGC TGT TTC ATC GAG GAA ATC CCC GAC GAG ACC ATG GTC ATC GGC AAC      145
Arg Cys Phe Ile Glu Glu Ile Pro Asp Glu Thr Met Val Ile Gly Asn
                 15                  20                  25

TAT CGT ACC CAG ATG TGG GAT AAG CAG AAG GAG GTC TTC CTG CCC TCG      193
Tyr Arg Thr Gln Met Trp Asp Lys Gln Lys Glu Val Phe Leu Pro Ser
             30                  35                  40

ACC CCT GGC CTG GGC ATG CAC GTG GAA GTG AAG GAC CCC GAC GGC AAG      241
Thr Pro Gly Leu Gly Met His Val Glu Val Lys Asp Pro Asp Gly Lys
         45                  50                  55

GTG GTG CTG TCC CGG CAG TAC GGC TCG GAG GGC CGC TTC ACG TTC ACC      289
Val Val Leu Ser Arg Gln Tyr Gly Ser Glu Gly Arg Phe Thr Phe Thr
     60                  65                  70

TCC CAC ABN KSG GGT GAC CAT CAA ATC TGT CTG CAC TGC GGG              331
Ser His Xaa Xaa Gly Asp His Gln Ile Cys Leu His Cys Gly
 75                  80                  85


(2) INFORMATION FOR SEQ ID NO: 180:

    (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 195 base pairs
          (B) TYPE: NUCLEIC ACID
          (C) STRANDEDNESS: DOUBLE
          (D) TOPOLOGY: LINEAR

    (ii) MOLECULE TYPE: CDNA

    (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Homo Sapiens
          (F) TISSUE TYPE: Normal prostate

    (ix) FEATURE:
          (A) NAME/KEY: other
          (B) LOCATION: 90..129
          (C) IDENTIFICATION METHOD: blastn
          (D) OTHER INFORMATION:  identity 100
                                  region 1..40
```

id AA134726
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 157..195
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 66..104
id AA134726
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 124..156
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 93
region 34..66
id AA134726
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 107..195
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 1..89
id R17226
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 76..138
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 12.7
seq ILFLLSWSGPLQG/QQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

```
AAGCTAACCC TCGGGCTTGA GGGGAAGAGG CTGACTGTAC GTTCCTTCTA CTCTGGCACC    60

ACTCTCCAGG CTGCC ATG GGG CCC AGC ACC CCT CTC CTC ATC TTG TTC CTT    111
                 Met Gly Pro Ser Thr Pro Leu Leu Ile Leu Phe Leu
                     -20                 -15                 -10

TTG TCA TGG TCG GGA CCC CTC CAA GGA CAG CAG CAC CAC CTT GTG GAG     159
Leu Ser Trp Ser Gly Pro Leu Gln Gly Gln Gln His His Leu Val Glu
                 -5                   1               5

TAC ATG GAA CGC CGA CTA GCT GCT TTA GAG GAA CGG                     195
Tyr Met Glu Arg Arg Leu Ala Ala Leu Glu Glu Arg
         10                  15
```

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 base pairs
(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 313..349
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 7..43
id T67245
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 119..199
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 8.8
seq LLLLCPLSRGCCP/LL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

```
ACGTTACCTT TGGGTGGTGG TTTTCATTCC TGTGCCGCCT GCTTCTGGGC CAGTGATCCA    60

GGTGTCTGGT GACCACCCGG GCACAGCTGC TTGGCTGCTG TGGGCACCTC AGCTTCCC     118

ATG TCC TGT AGG GAA CTC ACC CAC CGG CCT TGC TCT CCA CAC CTC TTA    166
Met Ser Cys Arg Glu Leu Thr His Arg Pro Cys Ser Pro His Leu Leu
    -25                 -20                 -15

CTC CTG TGT CCC CTT TCT CGG GGA TGC TGC CCC CTC CTG CTG TCC KGT    214
Leu Leu Cys Pro Leu Ser Arg Gly Cys Cys Pro Leu Leu Leu Ser Xaa
-10                 -5                   1                   5

CCY CTG TWA GGG GTG AAT CTT GAA TCC ATC TTA TCT CTT ACT CTC CCT    262
Pro Leu Xaa Gly Val Asn Leu Glu Ser Ile Leu Ser Leu Thr Leu Pro
                10                  15                  20

CCC TCT CCC AGC TCA GTC GGG CTC TCA CCC TCT GTG ACC CAS CTC ACA    310
Pro Ser Pro Ser Ser Val Gly Leu Ser Pro Ser Val Thr Xaa Leu Thr
            25                  30                  35

ACT TCA CCT GTT TCA TTG CAC TTT GCA TCC GMC CTC GCC GGG            352
Thr Ser Pro Val Ser Leu His Phe Ala Ser Xaa Leu Ala Gly
        40                  45                  50
```

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 113..306
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 99
region 71..264
id H83784
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 42..111
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 1..70
id H83784
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 378..414
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 94
region 346..382
id H83784
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 305..340
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 94
region 264..299
id H83784
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 250..350
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 2..102
id W32197
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 392..449
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 142..199
id W32197
est

```
(ix) FEATURE:
      (A) NAME/KEY: other
      (B) LOCATION: 349..390
      (C) IDENTIFICATION METHOD: blastn
      (D) OTHER INFORMATION:  identity 100
                              region 100..141
                              id W32197
                              est

(ix) FEATURE:
      (A) NAME/KEY: other
      (B) LOCATION: 397..449
      (C) IDENTIFICATION METHOD: blastn
      (D) OTHER INFORMATION:  identity 100
                              region 1..53
                              id W37255
                              est

(ix) FEATURE:
      (A) NAME/KEY: sig_peptide
      (B) LOCATION: 85..150
      (C) IDENTIFICATION METHOD: Von Heijne matrix
      (D) OTHER INFORMATION:  score 8.5
                              seq AALLLGLMMVVTG/DE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:


AACTTGTGTC CGGGTGGWRG ACTGGATTAG CTGCGGASCC TGGAAGCTGC CTGTCCTTCT    60

CCCTGTGCTT AACCAGAGGT GCCC ATG GGT TGG ACA ATG AGG CTG GTC ACA      111
                           Met Gly Trp Thr Met Arg Leu Val Thr
                                   -20                     -15

GCA GCA CTG TTA CTG GGT CTC ATG ATG GTG GTC ACT GGA GAC GAG GAT     159
Ala Ala Leu Leu Leu Gly Leu Met Met Val Val Thr Gly Asp Glu Asp
            -10                  -5                   1

GAG AAC AGC CCG TGT GCC CAT GAG GCC CTC TTG GAC GAG GAC ACC CTC     207
Glu Asn Ser Pro Cys Ala His Glu Ala Leu Leu Asp Glu Asp Thr Leu
     5                  10                  15

TTT TGC CAG GGC CTT GAA GTT TTC TAC CCA GAG TTG GGG AAC ATT GGC     255
Phe Cys Gln Gly Leu Glu Val Phe Tyr Pro Glu Leu Gly Asn Ile Gly
 20                 25                  30                  35

TGC AAG GTT GTT CCT GAT TGT DAC AAC TAC AGA CAG AAG ATC ACC TCC     303
Cys Lys Val Val Pro Asp Cys Xaa Asn Tyr Arg Gln Lys Ile Thr Ser
                40                  45                  50

TGG ATG GAG CCG ATA GTC AAG TTC CCG GGG GCC GTG GAC GGC GCA ACC     351
Trp Met Glu Pro Ile Val Lys Phe Pro Gly Ala Val Asp Gly Ala Thr
            55                  60                  65

TAT ATC CTG GTG ATG GTG GAT CCA GAT GCC CCT AGC AGA GCA GAA CCC     399
Tyr Ile Leu Val Met Val Asp Pro Asp Ala Pro Ser Arg Ala Glu Pro
        70                  75                  80

AGA CAG AGA TTC TGG AGA CAT TGG CTG GTA ACA GAT ATC AAG GGC GCC     447
Arg Gln Arg Phe Trp Arg His Trp Leu Val Thr Asp Ile Lys Gly Ala
    85                  90                  95
```

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 217 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 125..182
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 6..63
id R18560
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 176..213
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 58..95
id R18560
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 145..182
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 1..38
id R13864
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 176..213
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 33..70
id R13864
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 176..213
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 2..39
id HSC01E071
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 119..190
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.3
seq VHLLSLCSGKVYA/RM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

```
ACTGGGAGCC GCCTCCGTCG CCGCCGTCAG AGCCGCCCTA TCAGAGTTCC TACCANTTTG   60

TGGTTCCAGC AGCTTCTGTT CCAGATTATC TTAACAAGAA AACCAACTGG AAAAAAAA    118

ATG AAA TTC CTT ATC TTC GCA TTT TTC GGT GGT GTT CAC CTT TTA TCC    166
Met Lys Phe Leu Ile Phe Ala Phe Phe Gly Gly Val His Leu Leu Ser
                -20                 -15                 -10

CTG TGC TCT GGG AAA GTA TAT GCA AGA ATG GCA TCT CTA AGA GGA CTC    214
Leu Cys Ser Gly Lys Val Tyr Ala Arg Met Ala Ser Leu Arg Gly Leu
             -5                   1               5

GGG                                                                217
Gly
```

(2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 433 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 139..361
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 99
region 92..314
id AA100852
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 360..434
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 314..388
id AA100852
est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 139..434
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 100..395
id AA224847
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 139..361
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 92..314
id AA161042
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 368..434
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 323..389
id AA161042
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 139..365
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 87..313
id H64488
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 52..144
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 1..93
id H64488
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 171..396
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 129..354
id AA088770
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 167..253
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.1
seq LIFLCGAALLAVG/IW

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

```
AAAAAGCGCC TACCCTGCCT GCAGGTGAGC AGTGGTGTGT GAGAGCCAGG CGTCCCTCTG   60

CCTGCCCACT CAGTGGCAAC ACCCGGGAGC TGTTTTGTCC TTTGTGGAGC CTCAGCAGTT  120

CCCTCTTTCA GAACTYRVYK GCCAAGAGCC CTGAACAGGA GCCACC ATG CAG TGC     175
                                                   Met Gln Cys

TTC AGC TTC ATT AAG ACC ATG ATG ATC CTC TTC AAT TTG CTC ATC TTT    223
Phe Ser Phe Ile Lys Thr Met Met Ile Leu Phe Asn Leu Leu Ile Phe
    -25                 -20                 -15

CTG TGT GGT GCA GCC CTG TTG GCA GTG GGC ATC TGG GTG TCA ATC GAT    271
Leu Cys Gly Ala Ala Leu Leu Ala Val Gly Ile Trp Val Ser Ile Asp
-10                  -5                   1               5

GGG GCA TCC TTT CTG AAG ATC TTC GGG CCA CTG TCG TCC AGT GCC ATG    319
Gly Ala Ser Phe Leu Lys Ile Phe Gly Pro Leu Ser Ser Ser Ala Met
             10                  15                  20

CAG TTT GTC AAC GTG GGC TAC TTC CTC ATC GCA GCC GGC GTT GTG GTC    367
Gln Phe Val Asn Val Gly Tyr Phe Leu Ile Ala Ala Gly Val Val Val
        25                  30                  35

TTT GCT CTT GGT TTC CTG GGC TGC TAT GGT GCT AAG ACT GAG AGC AAG    415
Phe Ala Leu Gly Phe Leu Gly Cys Tyr Gly Ala Lys Thr Glu Ser Lys
     40                  45                  50

TGT GCC CTC GTG ACG TTC                                            433
Cys Ala Leu Val Thr Phe
 55                  60
```

(2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 128..242
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 1..115
id R58075
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 220..303

(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.6
seq IVSLLGFVATVTL/IP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

```
AAGATAGGCG GGTGCAGCGG GGCAGAACAT AGGTTGCCTT AGAGAGGTTC CCCGGAGTCC    60

CGACGGCGGC TCAAGTCAGA GTTGCTGGGT TTTGCTCAGA TTGGTGTGGG AAGAGCCTGC   120

CTGTGGGGAG CGGCCACTCC ATACTGCTGA GGCCTCAGGA CTGCTGCTCA GCTTGCCCGT   180

TACCTGAAGA GGCGGCGGAS GGGCCCCTGA CCGGTCACC ATG TGG GCC TTC TCG      234
                                          Met Trp Ala Phe Ser
                                                       -25

GAA TTG CCC ATG CCG CTG CTG ATC AAT TTG ATC GTC TCG CTG CTG GGA     282
Glu Leu Pro Met Pro Leu Leu Ile Asn Leu Ile Val Ser Leu Leu Gly
            -20                 -15                 -10

TTT GTG GCC ACA GTC ACC CTC ATC CCG GCC TTC CGG GGC CAC TTC ATT     330
Phe Val Ala Thr Val Thr Leu Ile Pro Ala Phe Arg Gly His Phe Ile
         -5                   1                   5

GCT GCG CGC CTC TGT GGT CAG GAC CTC AAC AAA ACC AGC CAG             372
Ala Ala Arg Leu Cys Gly Gln Asp Leu Asn Lys Thr Ser Gln
 10                  15                  20
```

(2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 402 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 112..403
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 33..324
id H97426
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 59..295
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 2..238
id W44834

est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 106..156
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 4..54
id R57989
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 161..190
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 93
region 62..91
id R57989
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 148..204
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 6.3
seq VLMRLVASAYSIA/QK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

```
AGCTGAGGTA GGGATGCSAT CCTTCTCAAA AGACTTATTG ACAGTGCCAA AGCTSGGTAC   60

TGGACACAAC GAGGGACCTG GGTCTACGAT AACGCGCTTK TGCTCCTCCT GAAGTGTCTT  120

TGGTCCAACG TTGTTCCAGA GTGTACC ATG GCT TCC AGT AAC ACT GTG TTG ATG  174
                              Met Ala Ser Ser Asn Thr Val Leu Met
                                              -15

CGG TTG GTA GCC TCC GCA TAT TCT ATT GCT CAA AAG GCA GGD ATG ATA     222
Arg Leu Val Ala Ser Ala Tyr Ser Ile Ala Gln Lys Ala Gly Met Ile
-10                 -5                  1                   5

GTC AGA CGT GTT ATT GCT GAA GGA GAC CTG GGT ATT GTG GAG ADG ACC     270
Val Arg Arg Val Ile Ala Glu Gly Asp Leu Gly Ile Val Glu Xaa Thr
            10                  15                  20

TGT GCA ACA GAC CTG CAG ACC AAA GCT GAC CGA TTG GCA CAG ATG AGN     318
Cys Ala Thr Asp Leu Gln Thr Lys Ala Asp Arg Leu Ala Gln Met Xaa
        25                  30                  35

ATA TGT TCT TCA TTG GCC CGG AAA TTC CCC AAA CTC ACA ATT ATA GGG     366
Ile Cys Ser Ser Leu Ala Arg Lys Phe Pro Lys Leu Thr Ile Ile Gly
    40                  45                  50

GAA GAG GAT CTG CCT TCT RMG GAA GTG GAT CAA GAG                     402
Glu Glu Asp Leu Pro Ser Xaa Glu Val Asp Gln Glu
 55                  60                  65
```

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 317 base pairs
    (B) TYPE: NUCLEIC ACID
    (C) STRANDEDNESS: DOUBLE
    (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Homo Sapiens
    (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
    (A) NAME/KEY: other
    (B) LOCATION: 111..318
    (C) IDENTIFICATION METHOD: blastn
    (D) OTHER INFORMATION: identity 99
        region 6..213
        id R18560
        est

(ix) FEATURE:
    (A) NAME/KEY: other
    (B) LOCATION: 131..318
    (C) IDENTIFICATION METHOD: blastn
    (D) OTHER INFORMATION: identity 99
        region 1..188
        id R13864
        est

(ix) FEATURE:
    (A) NAME/KEY: other
    (B) LOCATION: 162..318
    (C) IDENTIFICATION METHOD: blastn
    (D) OTHER INFORMATION: identity 99
        region 1..157
        id HSC01E071
        est

(ix) FEATURE:
    (A) NAME/KEY: other
    (B) LOCATION: 207..318
    (C) IDENTIFICATION METHOD: blastn
    (D) OTHER INFORMATION: identity 99
        region 1..112
        id AA016124
        est

(ix) FEATURE:
    (A) NAME/KEY: sig_peptide
    (B) LOCATION: 105..176
    (C) IDENTIFICATION METHOD: Von Heijne matrix
    (D) OTHER INFORMATION: score 5.9
        seq VHLLSLCSGKAIC/KN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

AAGTAAGGCT AGGYCGCGAG CTTAGTCCTG GGAGCCGCCT CCGTCGCCGC CGTCAGAGCC 60

```
GCCCTATCAG ATTATCTTAA CAAGAAAACC AACTGGAAAA AAAA ATG AAA TTC CTT   116
                                                 Met Lys Phe Leu

ATC TTC GCA TTT TTC GGT GGT GTT CAC CTT TTA TCC CTG TGC TCT GGG   164
Ile Phe Ala Phe Phe Gly Gly Val His Leu Leu Ser Leu Cys Ser Gly
-20                 -15                 -10                  -5

AAA GCT ATA TGC AAG AAT GGC ATC TCT AAG AGG ACT TTT GAA GAA ATA   212
Lys Ala Ile Cys Lys Asn Gly Ile Ser Lys Arg Thr Phe Glu Glu Ile
                  1               5                  10

AAA GAA GAA ATA GCC AGC TGT GGA GAT GTT GCT AAA GCA ATC ATC AAC   260
Lys Glu Glu Ile Ala Ser Cys Gly Asp Val Ala Lys Ala Ile Ile Asn
         15                  20                  25

CTA GCT GTT TAT GGT AAA GCC CAG AAC AGA TCC TAT GMG CGA TTG GCA   308
Leu Ala Val Tyr Gly Lys Ala Gln Asn Arg Ser Tyr Xaa Arg Leu Ala
     30                  35                  40

CTT CTG GTT                                                       317
Leu Leu Val
 45
```

(2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 499 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Normal prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 160..401
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 95
region 59..300
id H29377
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 454..499
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 100
region 356..401
id H29377
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 136..179
(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 95
region 36..79
id H29377
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 397..436
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 297..336
id H29377
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 135..295
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 95
region 293..453
id N28905
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 45..127
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 93
region 4..86
id N28905
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 334..388
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 94
region 489..543
id N28905
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 135..395
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 81..341
id H11885
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 160..384
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 95
region 105..329
id H15231
est

(ix) FEATURE:
(A) NAME/KEY: other

(B) LOCATION: 136..181
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 93
region 82..127
id H15231
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 146..298
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.9
seq ALXVLPLLGLHEA/AS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

```
AACTTCCGGG TTCGGCAATA ACCTGGAGCC GGCGGCGTAG GTTGGCTCTT TAGGGCTTCA   60

CCCCGAAGCT CCACCTTCGC TCCCGTCTTT CTGGAAACAC CGCTTTGATC TCGGCGGTGC  120

GGGACAGACG CTAGTGTGAG CCNMC ATG GCA GAT ACG ACC CCG AAC GGC CCC    172
                            Met Ala Asp Thr Thr Pro Asn Gly Pro
                                -50                 -45

CAA GGG GCG GGC GCT GTG CAA TTC ATG ATG ACC AAT AAA CTG GAC ACG    220
Gln Gly Ala Gly Ala Val Gln Phe Met Met Thr Asn Lys Leu Asp Thr
        -40                 -35                 -30

GCA ATG TGG CTT TCT CGC TTG TTC ACA GTT TAC TGC TCT GCT CTG NNT    268
Ala Met Trp Leu Ser Arg Leu Phe Thr Val Tyr Cys Ser Ala Leu Xaa
    -25                 -20                 -15

GTT CTG CCT CTT CTT GGG TTG CAT GAA GCA GCA AGC TTT TAC CAA CGT    316
Val Leu Pro Leu Leu Gly Leu His Glu Ala Ala Ser Phe Tyr Gln Arg
-10                  -5                   1                   5

GCT TTG CTG GCA AAT GCT CTT ACC AGT GCT CTG AGG CTG CAT CAA AGA    364
Ala Leu Leu Ala Asn Ala Leu Thr Ser Ala Leu Arg Leu His Gln Arg
             10                  15                  20

TTA CCA CAC TTC CAG TTA AGC AGA GCA TTC CTG GCC CAG GCT TTG TTA    412
Leu Pro His Phe Gln Leu Ser Arg Ala Phe Leu Ala Gln Ala Leu Leu
         25                  30                  35

GAG GAC AGC TGC CAC TAC CTG TTG TAT TCA CTC ATC TTT GTA AAT TCC    460
Glu Asp Ser Cys His Tyr Leu Leu Tyr Ser Leu Ile Phe Val Asn Ser
     40                  45                  50

TAT CCA GTT ACA ATG AGT ATC TTC CCA GTC TTG TTA TTC                499
Tyr Pro Val Thr Met Ser Ile Phe Pro Val Leu Leu Phe
 55                  60                  65
```

(2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 219 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 45..221
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 93
region 1..177
id HUMHBC4659
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 63..221
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 94
region 1..159
id AA160569
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 124..159
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 97..132
id R88362
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 1..72
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.5
seq XVLVLSVVXXAMA/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

```
ATG CGT TTC CGC CAT TTT TGM AAA TWA ATT GGG MAG GTA CTG GTT TTA      48
Met Arg Phe Arg His Phe Xaa Lys Xaa Ile Gly Xaa Val Leu Val Leu
                -20                 -15                 -10

AGT GTA GTT SCC GMC GCA ATG GCA GCC TTT GCA GTG SHA CCT CAG GGG      96
Ser Val Val Xaa Xaa Ala Met Ala Ala Phe Ala Val Xaa Pro Gln Gly
             -5                   1               5

CCC GCG TTA SSM TCT GAA CCA MTG MTG CYG GGT TCA CCC ACA TCT CCA     144
Pro Ala Leu Xaa Ser Glu Pro Xaa Xaa Xaa Gly Ser Pro Thr Ser Pro
     10                  15                  20

AAG CCA GGA GTT AAT GCC CAG TTC TTA CCT GGA TTT TTA ATG GGG GMT     192
Lys Pro Gly Val Asn Ala Gln Phe Leu Pro Gly Phe Leu Met Gly Xaa
 25                  30                  35                  40
```

```
TTG CCA GCT CCG GTG ACT CCA CAA CCT                                219
Leu Pro Ala Pro Val Thr Pro Gln Pro
                45
```

(2) INFORMATION FOR SEQ ID NO: 190:

```
    (i) SEQUENCE CHARACTERISTICS:
         (A) LENGTH: 483 base pairs
         (B) TYPE: NUCLEIC ACID
         (C) STRANDEDNESS: DOUBLE
         (D) TOPOLOGY: LINEAR

    (ii) MOLECULE TYPE: CDNA

    (vi) ORIGINAL SOURCE:
         (A) ORGANISM: Homo Sapiens
         (F) TISSUE TYPE: Normal prostate

    (ix) FEATURE:
         (A) NAME/KEY: other
         (B) LOCATION: 105..414
         (C) IDENTIFICATION METHOD: blastn
         (D) OTHER INFORMATION:  identity 99
                                 region 1..310
                                 id T26956
                                 est

    (ix) FEATURE:
         (A) NAME/KEY: other
         (B) LOCATION: 45..359
         (C) IDENTIFICATION METHOD: blastn
         (D) OTHER INFORMATION:  identity 98
                                 region 1..315
                                 id T31666
                                 est

    (ix) FEATURE:
         (A) NAME/KEY: other
         (B) LOCATION: 202..332
         (C) IDENTIFICATION METHOD: blastn
         (D) OTHER INFORMATION:  identity 100
                                 region 137..267
                                 id R14990
                                 est

    (ix) FEATURE:
         (A) NAME/KEY: other
         (B) LOCATION: 127..201
         (C) IDENTIFICATION METHOD: blastn
         (D) OTHER INFORMATION:  identity 100
                                 region 63..137
                                 id R14990
                                 est

    (ix) FEATURE:
         (A) NAME/KEY: other
         (B) LOCATION: 65..114
         (C) IDENTIFICATION METHOD: blastn
         (D) OTHER INFORMATION:  identity 100
```

```
                                   region 1..50
                                   id R14990
                                   est

     (ix) FEATURE:
          (A) NAME/KEY: sig_peptide
          (B) LOCATION: 1..120
          (C) IDENTIFICATION METHOD: Von Heijne matrix
          (D) OTHER INFORMATION:  score 5.2
                                  seq LCVEFASVASCDA/AV

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

ATG GAG TTG GGG AGT TGC CTG GAG GGC GGG AGG GAG GCG GCG GAG GAA      48
Met Glu Leu Gly Ser Cys Leu Glu Gly Gly Arg Glu Ala Ala Glu Glu
-40                 -35                 -30                 -25

GAG GGC GAG CCT GAG GTG AAA AAG CGG CGA CTT CTG TGT GTG GAG TTT      96
Glu Gly Glu Pro Glu Val Lys Lys Arg Arg Leu Leu Cys Val Glu Phe
                -20                 -15                 -10

GCC TCG GTC GCA AGC TGC GAT GCC GCA GTG GCT CAG TGC TTC CTG GCC     144
Ala Ser Val Ala Ser Cys Asp Ala Ala Val Ala Gln Cys Phe Leu Ala
             -5                   1               5

GAG AAC GAC TGG GAG ATG GAA AGG GCT CTG AAC TCC TAC TTC GAG CCT     192
Glu Asn Asp Trp Glu Met Glu Arg Ala Leu Asn Ser Tyr Phe Glu Pro
     10                  15                  20

CCG GTG GAG GAG AGC GCC TTG GAA CGC CGA CCT GAA ACC ATC TCT GAG     240
Pro Val Glu Glu Ser Ala Leu Glu Arg Arg Pro Glu Thr Ile Ser Glu
 25                  30                  35                  40

CCC AAG ACC TAT GTT GAC CTA ACC AAT GAA GAA ACA ACT GAT TCC ACC     288
Pro Lys Thr Tyr Val Asp Leu Thr Asn Glu Glu Thr Thr Asp Ser Thr
                 45                  50                  55

ACT TCT AAA ATC AGC CCA TCT GAA GAT ACT CAG CAA GAA AAT GGC AGC     336
Thr Ser Lys Ile Ser Pro Ser Glu Asp Thr Gln Gln Glu Asn Gly Ser
             60                  65                  70

ATG TTC TCT CTC ATT ACC TGG AAT ATT GAT GGA TTA GAT CTA AAC AAT     384
Met Phe Ser Leu Ile Thr Trp Asn Ile Asp Gly Leu Asp Leu Asn Asn
         75                  80                  85

CTG TCA GAG AGG GCT CGA GGG GTG TGT TCC TAC TTA GCT TTG TAC AGC     432
Leu Ser Glu Arg Ala Arg Gly Val Cys Ser Tyr Leu Ala Leu Tyr Ser
     90                  95                 100

CCA GAT GTG ATA TTT CTA CAG GAA GTT ATT CCC CCA TAT TAT AGC TAC     480
Pro Asp Val Ile Phe Leu Gln Glu Val Ile Pro Pro Tyr Tyr Ser Tyr
105                 110                 115                 120

CTA
Leu                                                                 483

(2) INFORMATION FOR SEQ ID NO: 191:
```

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 182..401
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 165..384
id W56608
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 45..130
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 30..115
id W56608
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 127..191
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 96
region 111..175
id W56608
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 401..446
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 385..430
id W56608
est

(ix) FEATURE:
(A) NAME/KEY: other
(B) LOCATION: 311..446
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 99
region 1..136
id R17248
est

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 13..378
(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 5
seq RLVVVSVSPQSRA/SL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

```
AGTGCGGCCG TC ATG GCG TCG CCC TTC AGC GGG GCG CTG CAG CTG ACG GAC    51
              Met Ala Ser Pro Phe Ser Gly Ala Leu Gln Leu Thr Asp
                      -120                -115                -110

CTG GAT GAC TTC ATC GGG CCG TCT CAG GAG TGC ATC AAG CCT GTC AAA      99
Leu Asp Asp Phe Ile Gly Pro Ser Gln Glu Cys Ile Lys Pro Val Lys
            -105                -100                -95

GTG GAA AAA AGG GCG GGA AGT GGC GTG GCC AAG ATT CGC ATT GAA GAT     147
Val Glu Lys Arg Ala Gly Ser Gly Val Ala Lys Ile Arg Ile Glu Asp
            -90                 -85                 -80

GAC GGG AGC TAC TTC CAA ATT AAC CAA GAC GGC DGG ACC CGG AGG CTG     195
Asp Gly Ser Tyr Phe Gln Ile Asn Gln Asp Gly Xaa Thr Arg Arg Leu
        -75                 -70                 -65

GAG AAG GCC AAG GTC TCG CTA AAC TAC TGC NWG GCG TGC AGC GGC TGC     243
Glu Lys Ala Lys Val Ser Leu Asn Tyr Cys Xaa Ala Cys Ser Gly Cys
    -60                 -55                 -50

ATC ACC TCC GCA GAG ACC GTG CTT ATC ACC CAG CAG AGC CAC GAG GAG     291
Ile Thr Ser Ala Glu Thr Val Leu Ile Thr Gln Gln Ser His Glu Glu
-45                 -40                 -35                 -30

CTG AAG AAG GTT CTA GAT GCT AAC AAG ATG GCG GCA CCC AGT CAG CAG     339
Leu Lys Lys Val Leu Asp Ala Asn Lys Met Ala Ala Pro Ser Gln Gln
                -25                 -20                 -15

AGG CTG GTT GTA GTT TCG GTC TCA CCA CAG TCT AGA GCA TCG CTG GCT     387
Arg Leu Val Val Val Ser Val Ser Pro Gln Ser Arg Ala Ser Leu Ala
            -10                 -5                  1

GCA CGG TTT CAG CTG AAW CCT ACA GAT ACT GCC AGG AAA TTA ACC TCA     435
Ala Arg Phe Gln Leu Xaa Pro Thr Asp Thr Ala Arg Lys Leu Thr Ser
    5                   10                  15

TTC TTT AAA                                                         444
Phe Phe Lys
20
```

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 335 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: Prostate

THIS PAGE BLANK (USPTO)